

-251-

Query: 63 D---NNSKIA 69
 D N+S IA
 Sbjct: 62 DTTENDSLIA 71

5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 581> which encodes the amino acid sequence <SEQ ID 582>. Analysis of this protein sequence reveals the following:

Possible site: 30
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0680 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15 An alignment of the GAS and GBS proteins is shown below:

Identities = 21/61 (34%), Positives = 34/61 (55%)

Query: 1 MYNRLKELRKDKGLTQADLAKVINTNQSQYGYENGKTSLSIENSKILADFFGVSIPLL 60
 MY R++ LR+D TQ +A +++ + + Y K E G+ +L + + VSI YLL
 20 Sbjct: 1 MYPRIRNLREDNDFTQKFVANLLSFSHANYAKIERGEVALMADVILVQFYKLYNVSIDVLL 60
 Query: 61 G 61
 G
 Sbjct: 61 G 61

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 174

A DNA sequence (GBSx0180) was identified in *S.agalactiae* <SEQ ID 583> which encodes the amino acid sequence <SEQ ID 584>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 35 bacterial cytoplasm --- Certainty=0.5278 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

40 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 175

A DNA sequence (GBSx0181) was identified in *S.agalactiae* <SEQ ID 585> which encodes the amino acid sequence <SEQ ID 586>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.3762 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 176

A DNA sequence (GBSx0182) was identified in *S.agalactiae* <SEQ ID 587> which encodes the amino acid sequence <SEQ ID 588>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
10      INTEGRAL    Likelihood = -9.66    Transmembrane   40 - 56 ( 33 - 65)
      INTEGRAL    Likelihood = -5.79    Transmembrane   62 - 78 ( 59 - 81)

----- Final Results -----
15      bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8505> and protein <SEQ ID 8506> were also identified. Analysis of this 20 protein sequence reveals the following:

```
Lipop: Possible site: -1 Crend: 7
McG: Discrim Score: -16.96
GvH: Signal Score (-7.5): -2.95
      Possible site: 57
25      >>> Seems to have no N-terminal signal sequence
      ALOM program count: 2 value: -9.66 threshold: 0.0
      INTEGRAL    Likelihood = -9.66    Transmembrane   33 - 49 ( 26 - 58)
      INTEGRAL    Likelihood = -5.79    Transmembrane   55 - 71 ( 52 - 74)
      PERIPHERAL  Likelihood = 10.87    14
30      modified ALOM score: 2.43

*** Reasoning Step: 3

----- Final Results -----
35      bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 177

A DNA sequence (GBSx0183) was identified in *S.agalactiae* <SEQ ID 589> which encodes the amino acid sequence <SEQ ID 590>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
45      >>> Final Results -----
      bacterial cytoplasm --- Certainty=0.3276 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

50 The protein has no significant homology with any sequences in the GENPEPT database.

-253-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 178

- 5 A DNA sequence (GBSx0184) was identified in *S.agalactiae* <SEQ ID 591> which encodes the amino acid sequence <SEQ ID 592>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
```

```
10 ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3482 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

- 15 A related GBS nucleic acid sequence <SEQ ID 9509> which encodes amino acid sequence <SEQ ID 9510> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA30291 GB:X07371 RepM protein (AA 1 - 314) [Staphylococcus
aureus]
20 Identities = 89/283 (31%), Positives = 145/283 (50%), Gaps = 26/283 (9%)
Query: 67 KVSLDNITMTAYIKSKKYLAMQQLIETHLAITVQTAMTDMFRATTGDGIHVVLHMNYDKQ 126
      K+S D +T+ + + + I + + F+A + +++ YDK
Sbjct: 42 KLSFDAMTIVGNLNKNSAKKLSDFMSLDPQIRLWDILQTKFKAKA---LQEKVYIEYDKV 98
25 Query: 127 KGQDRKARPFRLEFPNPNKRLRVDSEII---DTIIPFLEDISISRADLAFLFEVDCSEF- 182
      K R R+EFNPNKL E++ II ++ED +R DLAFL FE D S++
Sbjct: 99 KADTWDRRNMRVFPNPK---THDEMLWLKHNIIDYMEDDGFTRLDRAFT-FEDDLSDYY 155
30 Query: 183 -VLEKKGRPTATKEFRSSTGTLETKYLGAPRSEKQVRLYNNKKEQLQNGTDKDGFASQF 241
      + EK + T F +TG ETKY G+ S + +R+YNKKKE+ +N D D +++
Sbjct: 156 ALSEKALKRTV---FFGTTGKAETKYFGSRDSNRFIRIYNKKKERKENA---DVDVSAE- 208
Query: 242 KHWWRLEFQLRSRSIDEIFEVI-DTIIFKP--FNLKGLSIETQIYLTALIHDKNIWKKLH 298
      H WR+E +L+ +D D I KP L+ L + +YL L+H+++ W +LH
Sbjct: 209 -HLWRVEIELKRDMDYWNNCFNDLHILKPAWATLESLKEQAMVYL--LLHEESKWGELH 265
35 Query: 299 RNTRARYKKILETHQTSDTDYLGLLKDLLKHERPRLENQLAYY 341
      RN+R +YK+I++ + S D L+K L L+ Q+ ++
Sbjct: 266 RNSRRKYKQIIQ--EISSIDLTMKSTLDNEENLQKQINFW 306
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 179

- A DNA sequence (GBSx0185) was identified in *S.agalactiae* <SEQ ID 593> which encodes the amino acid sequence <SEQ ID 594>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
50 INTEGRAL Likelihood =-15.55 Transmembrane 137 - 153 ( 133 - 157)
----- Final Results -----
    bacterial membrane --- Certainty=0.7220 (Affirmative) < succ>
```

-254-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9511> which encodes amino acid sequence <SEQ ID 9512>

5 was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8507> and protein <SEQ ID 8508> were also identified. Analysis of this protein sequence reveals the following:

```

10    Lipop: Possible site: -1 Crend: 2
      McG: Discrim Score: -16.84
      GvH: Signal Score (-7.5): -5.3
            Possible site: 32
      >>> Seems to have no N-terminal signal sequence
15    ALOM program count: 1 value: -15.55 threshold: 0.0
          INTEGRAL Likelihood =-15.55 Transmembrane 137 - 153 ( 133 - 157)
          PERIPHERAL Likelihood = 10.93      60
      modified ALOM score: 3.61

20    *** Reasoning Step: 3

----- Final Results -----
      bacterial membrane --- Certainty=0.7220(Affirmative) < suc>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

	ORF01844 (292 - 702 of 1074)							
30	EGAD 124517 132830(149 - 295 of 435) apolipoprotein A-IV {Mus musculus}							
	GP 191889 gb AAA37216.1 M64250 apolipoprotein A-IV {Mus musculus castaneus}							
	%Match = 4.6							
	%Identity = 30.0 %Similarity = 54.6							
	Matches = 39 Mismatches = 53 Conservative Sub.s = 32							
35	201 231 261 291 321 351 381 411							
	NNSNIRY* LFRFAERLVEA *KTKTRKSARLLWG* DRQK* LSTLLLKIQYYQVQVTRSKMRRIKYADSLGVSSQSITYKRIRS							
	LRDRMMMPHANKVTQTFGENMQKLQEHLKPYAVDLQDQINTQTQEMKLQLTPYIQRMQTTIKENVDNLHTSMMPLATNLKD							
	120 130 140 150 160 170 180							
40	435 462 492 522 552 570							
	P--KYKERLKGHLY-RDNQKVENVLDLIGIKILEDYHFENDVIELEKTLDG-----IQEEFEQEKKGMQY-----							
	: : : :: : : : : : : : : : : : : : :							
	KFNRRNMEELKGHLTPRANELKATID-----QNLEDLRSLSAPLTGVQEKLHNQMEGLAFQMKKNAAEELQTK							
45	200 210 220 230 240 250							
	615 645 672 702 732 762 792 822							
	-- RIDRLADKLTLPLIEDNQNLUVQKNYE-LLNYVRSLEROKLLLIIALAVMVITLLVAIWLAI F*WQLSDNAKRPTKGTA							
	: : : : : : : :							
50	VSAKIDQLQKNLAPLVEDVQSKVKGNTTEGLOKSILKDLNRQLEQQVEEFRRTVEPMGEMFNKALVQQLEQFRQQLGPNSGE							
	270 280 290 300 310 320 330							

SEQ ID 8508 (GBS405) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 171 (lane 4; MW 46kDa – 2 bands) and in Figure 177 (lane 7; MW 46kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 76 (lane 5; MW 21kDa).

GBS405-GST was purified as shown in Figure 218, lane 8.

-255-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 180

A DNA sequence (GBSx0186) was identified in *S.agalactiae* <SEQ ID 595> which encodes the amino acid sequence <SEQ ID 596>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.3406 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
15 >GP:CAA33713 GB:X15669 pre protein (AA 1-494) [Streptococcus
      agalactiae]
      Identities = 171/402 (42%), Positives = 250/402 (61%), Gaps = 46/402 (11%)

20 Query: 1 MSYVVARMAYKSGQLTAYNHNERIFKNHSNKEIDVEKSHLNYELTNRDQAQNYHKQIK 60
      MSY+VARM K K+G L + HNER+F+ HSNK+I+ +SHLNYELT+RD++ +Y KQIK
      Sbjct: 1 MSYMVARMQMKAGNLGGAFKHNERVFETHSNKDINPSRSHLNYELTDRDRSVSYEKQIK 60

25 Query: 61 EHINENRLSTRGVRKDAILCNEWIITSKTDKTFDSLDEKQTREFFETAKDYFAEKYGDANI 120
      +++NEN++S R +RKDA+LC+EWIITSKDF FF+ LDE+QTR FFETAK+YFAE YG++NI
      Sbjct: 61 DYVNENKVSNRAIRKDAVLCDEWIITSKDFEKLDEEQTRFFETAKNYFAENYGESNI 120

30 Query: 121 AYARVHLDESTPHMHLGIVPMKNGKLSSKALFGNKEKLVAIQDELPKYLNEHGFNLQRGE 180
      AYA VHLDESTPHMH+G+VP +NGKLSSKA+F ++E+L IQ++LP+Y+++HGF L+RG+
      Sbjct: 121 AYASVHLDESTPHMHMGVVPFENGKLSSKAMF-DREELKHIQEDLPRYMSDHGFELERGK 179

35 Query: 181 IGSKKKHEATAEFKEKQRLLDNADRKLADKHEELKALDDKISNV-NDTIA----- 229
      + S+ KH AEEFK ++ +L +K+ +D++ + NDT A
      Sbjct: 180 LNSEAKHKTVAAEFKRAMADME-LKEELLEKYHAPPFDERTGELNNNDTEAFWHEKEFADM 238

40 Query: 230 -DKESRLKEL---EAKEWDAGDLKQYELEKQSLAESIEDIKDIELLQLDRIQKEDILVKQ 285
      + +S ++E E +W KQY+ E + L S + ++D D E+L+ +
      Sbjct: 239 FEVQSPIRETTNQEKMDSLWRL---KQYQEELKKLESSKKPLED-----DLSHLEELLDK 288

45 Query: 286 SFDGKLKMDKETYNRLFQTASKHASSNAELKRDVLVKAQSQQNNHLSRELLNHRKTAEKNIK 345
      +K+D E AS+ AS +L KA+ N L NH K+ E I+
      Sbjct: 289 KTKEYIKIDSE-----ASERAS-----ELSKAEGYINTLE---NHSKSLEAKIE 329

Query: 346 LSQENRKLKDVKVMLDEQVKILNKSLSVWKEAKEFMPKQVY 387
      + + +K K + K LN+S + K F+ K+ Y
Sbjct: 330 CLESNDLQLEKQKATKLEAKALNESELRELKPKNFLGKEHY 371
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 597> which encodes the amino acid sequence <SEQ ID 598>. Analysis of this protein sequence reveals the following:

```
50 LPXTG motif: 2025-2030

      Possible site: 52

      >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood =-10.08 Transmembrane 2034 -2050 (2030 -2053)
      INTEGRAL Likelihood = -6.05 Transmembrane 21 - 37 ( 20 - 39)

----- Final Results -----
      bacterial membrane --- Certainty=0.5034 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

>GP:AAD03320 GB:AF067776 extracellular matrix binding protein
[Abiotrophia defectiva]

5 Identities = 362/1396 (25%), Positives = 591/1396 (41%), Gaps = 87/1396 (6%)

Query: 636 KAEVKLKEAHEATKQAIKEKDPWLSPEQKKAQKEKAKARLDEGLKALKAAADSLEILKVTEE 695
+A+ + A +A AI+ + L+ E+K A+K K +A + L + A K T

10 Sbjct: 636 EAKNAVNNAAKAKNTAIDNNNNLTAEKAAEAKVAAKNATLAGIDQA-----KTTAA 689

Query: 696 AFVDKEKNPDSIPNQHKAGTADQARKQALDSLKEVQKELESIDNDNTLTDEKAAKKK 755
+ K I + A A AL+ + ++ I LT +EK A +

Sbjct: 690 RNAAQNKGTTDINAVNPVPVAKPAANAALE---QAAVNKINEISQRPDLTREEKQAFMDQ 746

15 Query: 756 VNDAYDVAKQTAMEANSYEDLTTIKDEFLLS---NLPHKQGTPPLKDQQSDAIAELEKKQQE 812
V A D A A + + +T+ +D+ L+ NLP TP + +A+ + +
Sbjct: 747 VRTARDAAMAKVASAANNOQAVTSARDQGLNAVNNLP---TPAA-KYPEALGHVRQAADA 801

20 Query: 813 IEKAIEGDKTLPRDEKEKQIADSKERLKSDTQKVDAKNADAIIKKAFEEGKVNIPOQAHIP 872
+AI + L +E+ + + + + + KA +G I
Sbjct: 802 KRQAIRDNANLTAAEQADALRQVDAATQAAEAAINQNHTNATLAKADSDGVKAI----- 855

25 Query: 873 GDILN---KDKEKLLAELKQKADDTEKAIDVDKTLTEDEKKEQKVTKAELEKAKTDVKNT 929
D+N + K L+Q A +AI+ + LT++EK + + L AKT V+
Sbjct: 856 NDINPQPRSKPAANQALEQVAAKRQAINNNNQLTDEEKAQAIQQVDQALANAKTQVQAA 915

30 Query: 930 QTREELDKKVPVELKKAIEDTHVKGNLEGVKNKAIEDLKKAHTETWAKINGDDTLKDATKE 989
+++ AI + +G K +AI + + A + + + G + L +
Sbjct: 916 NDNNGVNQAKTAGTTAINNINPQGTQ---KAQATAAIEAAEQAKRLELQGRNDLTTEERN 972

Query: 990 AQVKEADKALAAGKDAITKADDADKVSTAVTEHTPKIKAHKTGDLKKAQVDANTALDKA 1049
+ + A KDA+ +A + V+ A +I+ + T +K DA A+D+A
Sbjct: 973 NALADLTAKAQAQAKDAVNQARNNTGVAQAKDNGVAQIQGINPTAVVKP---DARNAIDQA 1029

35 Query: 1050 AEKERGEINKDATLTTEDKAKQKVEVETALTKAKDNVKAAKTADAINDARDKGVTIDAV 1109
A + E + LT E+KA +K+V+ A AK + A + +N+A ++G A I A+
Sbjct: 1030 ARDKEAEFQANTKLTDEEKAIAKVKQDAARDAKAAIDRAGSNGDVNNNAVQGKAAIQAI 1089

40 Query: 1110 HKAGQDLGARKSGQVAKLEEAQAKATKDKISADPTLTSKEKEEQSKAVDAELKKAIEAVNA 1169
+ K A ++ AA A K I+A+ LT +EK K V+ E KA AV+A
Sbjct: 1090 KALDDSQPSAKDTAKAAIQNAADAKKAITANNALTQEKEAAAIIKQVEDEAAKAQAAVDA 1149

45 Query: 1170 ADTADKVDDALGEGVTDIKNQHKSGDSIDARREAHGKELDRVAQFTKGATEKDPTLTTEE 1229
+ + VD A +G+ I + ++ + +D+ A + K I D TLT EE
Sbjct: 1150 SRSKADVDRAKDQGLQKISDV---PAVQPPKLNAAIAAVDQAATDKKAVINNDTLLTQEE 1205

50 Query: 1230 KAKQVKDVDAAKERGMALKNEAKDADALDKAYGEGVTDIKNQHKSGDPVDARRGLHNKSI 1289
K ++ VD + +N+A + + +G I N ++ A + ++
Sbjct: 1206 KEAAIRKVDEEAAKARQAINDATSNADVAAKQAOQGTQAINNVPQT---PAAKNAAKAAV 1261

Query: 1290 DEVAQATKDAITADTTLTEAEKETQRGNVDKEATKAKEELAKAKDADALDKAYGDGVTSI 1349
++ A A K AI D LT EK+ VD+E KA++ + A + +G +I
Sbjct: 1262 EQAADAKKQAIENDPNLTRQEKDAIAKVDQETNKARQAIADAATTNADVTAKQNEGQTQAI 1321

55 Query: 1350 KNQHKSGKGKLDVRKDEHKKALEAVAKRVTAETEADPTLTPEVREQQKAEVQKELELATDK 1409
++ K K + K A+ A+ + IE DP LT E ++ KA+V E A +
Sbjct: 1322 NAVPQTPKA---KTDANKAVTQAAEDKKSAIENDPNLTREEKDAAKAKVDAEATKAKNA 1377

60 Query: 1410 IAEAKDADEADKAYGDGVTAIENAHVGKIGEARKDLAKKDLAEAAAKTKALIIEDKTLT 1469
I A D+ +G AI + + + +A+ D AK + +AA + K I D LT
Sbjct: 1378 IDAATSNDDETAKQNEGQTQAI---NAVQTPKAKTD-AKNAVTQAAADRKKDAIENDPNLT 1433

65 Query: 1470 DDQRKEQLLGVDTEYAKGIENIDAQDAAGVDKAYSQDGVRDILAQYKEGQNLNDRRNAAK 1529
+++ VD E K + IDAA A V ++G + I + + + AK
Sbjct: 1434 REEKVAAKAKVDAEAKKAKDAIDAATSNAADVTAQNEGQTQAI---NDVPQTPTAKTDK 1489

Query: 1530 EFLLKEADKVTKLINDDPTLTHDQKVQINKVEQAKLDAIKSVDDAQTADAINDALGKGI 1589

-257-

- + + AD I DP LT ++K KV+ A ++D A + + + G
- Sbjct: 1490 NAVTQAADAKKDAIEKDPNLTREEKDAAKAKVDAEAKKAKDAIDAATSNAADVTAQNEG 1549
- 5 Query: 1590 ENINNQYQHGDGVDRKATAKGDLKEAAKVKALIAKDP TL TQADKDQTA AVDAAKNTA 1649
+ IN+ Q K AK + + A K I KDP LT+ +KD A VDA A
- Sbjct: 1550 KAINDVPQ---TPTAKTDKNAVTQAADAKKDAIEKDPNLTREEKDAAKAKVDAEAKKA 1605
- 10 Query: 1650 IAAVDKATTTEGINQELGKGITAINKAYRPGEVKARKEAAKADLEKEAAKVKALITNDP 1709
A+D AT+ + + G AIN + K AK + + A K I ND
- Sbjct: 1606 KDAIDAATSNAADVTAQKDAGKNAINAVPQ---TPTAKTDKNAVTQAADAKKDAIENDA 1661
- 15 Query: 1710 TLTAKD-KAQTEAVAKALKAAIAAVDKATTAEGINQELGKGITAINKAYRPGEVKARK 1768
LT+ +K A + + A+A KA A+D AT+ + + +G AIN + K
- Sbjct: 1662 NLTREEKDAAKAKVDAEATKAK-NAIDAATSNAADVTAQNEGTKAINDVPQ---TPTAK 1716
- 20 Query: 1769 EAAKADLEREAAKVREAIANDPTLTAKD-KAQTEAVAKALKAAIAAVDKATTAEGINQE 1827
AK + + + A + AI NDP LT+ +K A + + A+A KA A+D AT+ + +
- Sbjct: 1717 TDAKNAVDQAA TDKSAIENDPALTREEKDAAKAKVDAEATKAK-NAIDAATSNAADVTAQ 1775
- 25 Query: 1888 ALAKALGQVEAAKTVEAVKLAENLGTVAIRSAYVAGLAKDTDQATAALNEAKQAAIEALK 1947
KA + +AA + V + + G KD A AK A A+
- Sbjct: 1832 EAKKAKDAIDAATSNAADVTAQKDAG-----KDAINAVPQTPTAKTDKNAVD 1878
- 30 Query: 1948 QAAAETLAKITTDAKLTEAQKAEQSENVS LALKTAIATVRSAQSIAVSKEAKDKGITAIR 2007
QAA + + I D LT +K V K A + +A S A V + +G AI
- Sbjct: 1879 QAATDKKSAIENDPALTREEKDAVKA KVDAEAKKAKDAIDAATSNAADVTAQTEGTQAIN 1938
- Query: 2008 AAYVPNKAVAKSSSAN 2023
A VP AK+ + N
- 35 Sbjct: 1939 A--VPQTPTAKTDKNAV 1952
- An alignment of the GAS and GBS proteins is shown below:
- Identities = 77/396 (19%), Positives = 157/396 (39%), Gaps = 48/396 (12%)
- 40 Query: 42 LNYELTNRDQAQONYHKQI KEHINENRLSTRGVRKDAILCNEWIITS DKTFFDSLDEKQTR 101
L++E+ + ++QN K+I + + D E +I K + + + EK T
- Sbjct: 338 LD FEILH-PRSQNVSKKISKQVEAKPF-----DPASYKEKVIAKLKPVYEATSEKITN 389
- 45 Query: 102 EFF--ETAKDYFAEKYGDANIA YARVHLDESTPHMHIGIVPMKNGKLSSKALFG--NKEK 157
+ + E AKD +K + I+ G V + +A+ NK
- Sbjct: 390 DAWLDENAKDLQKQKLEEQYIS-----GKVAISEAGTKQEAIDAA YNKYS 434
- 50 Query: 158 LVAIQDEL PKYLNEHGFNLQRGEIGSKKKHLETAEFKEKQRLDN---ADRKLADKHEEL 214
D LP + N + + ++ + +T + K D K K E L
- Sbjct: 435 SQTDPDSLPSQYKQG--NKENEQEKG RQDLIQTRDLTLKAIQEDKWLTEQEKTIQKEEAL 492
- Query: 215 KALDDKISNVNDTIADKESRLKELEAKEWDAVGDLKQYE-----LEKOSLAESIE 264
KA + I +VN T++ + + + K + + K+Y EK+ A E
- Sbjct: 493 KAFETGIESVNQTVSLEQLKQRLIVYKASEKDSEKKEYPESIPNQHIPGKEKEVKAQK 552
- 55 Query: 265 DIKDIELLQLDRIQKEDLVKQSF DGKLKMDKETYNR LFQTA SKHASSMAELKRD LVKAQS 324
++K + L++I ++ + + E + Q A K A + +L+ DL S
- Sbjct: 553 ELKKLHD TTLEKINQDKWLTPDQQAEQLKQAEVTFKKGQEAIKSAQTLQLETDLAD YVS 612
- 60 Query: 325 QNNHLSRELLNHRKTAEKNIKLSQENRKLKDVKMLDEQVK---ILMKSLSVWKEAKE 380
+N + + K+ K+ + + + KLK+ + + ++ + + KEKAK
- Sbjct: 613 ENEGKGNSIPDKYKSGNKDDLVNKAEVKLKEAHEATKQAEKDPWLSPEQKKAQKEKAKA 672
- 65 Query: 381 FMPKQVYRETLSIINTLNPIGLAKTAIRQVKMVDS 416
+ + + + L ++L + + + A +K DS
- Sbjct: 673 RLDEGL--KALKAADSLEILKVTEEAFVDKEKNPDS 706

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 181

A DNA sequence (GBSx0187) was identified in *S.agalactiae* <SEQ ID 599> which encodes the amino acid sequence <SEQ ID 600>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2544 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 182

A DNA sequence (GBSx0188) was identified in *S.agalactiae* <SEQ ID 601> which encodes the amino acid sequence <SEQ ID 602>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25 bacterial cytoplasm --- Certainty=0.2045 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 603> which encodes the amino acid sequence <SEQ ID 604>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35 bacterial cytoplasm --- Certainty=0.2045 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

40 An alignment of the GAS and GBS proteins is shown below:

Identities = 102/111 (91%), Positives = 107/111 (95%)

```
Query: 1 MDYKKYQIIYAPDVLEKLKEIRDYISQNYSSTSGQHKMEOIIISDIEKLEVFPFVGFDADE 60
+DYKKYQIIYAPDVLEKLKEIRDYISQNYSSTSGQ KMEQIISDIEKLEVFPFVGFDADE
```

45 Sbjct: 1 LDYKKYQIIYAPDVLEKLKEIRDYISQNYSSTSGQRKMEQIISDIEKLEVFPFVGFDADE 60

```
Query: 61 KYGSKISKYHSTRGYTLSKDYIVLYHIEEEEENRVRVIDYLPLPTRSDYMKLFK 111
KYGSKI YHST+GYTLSKDYIVLYHIE EENR+VIDYLLPT+SDY+KLFK
```

Sbjct: 61 KYGSKIIHYHSTKGYTLSKDYIVLYHIEGEENRIVIDYLPTQSDYIKLFK 111

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 183

A DNA sequence (GBSx0189) was identified in *S.agalactiae* <SEQ ID 605> which encodes the amino acid sequence <SEQ ID 606>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.1621(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 607> which encodes the amino acid sequence <SEQ ID 608>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.1596(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

25 An alignment of the GAS and GBS proteins is shown below:

```

Identities = 91/95 (95%), Positives = 93/95 (97%)

Query: 1 MVTAEKNRARTFQANKELVSEAMTVLNKKNLTLSALRLFLQNVVVTNEVDLLTEEELEK 60
       M T +KNRAVTFQANKELVSEAMTVLNKKNLTLSALRLFLQNVVVTNEVDLLTEEELEK
30     Sbjct: 1 MTTVKKNRAVTFQANKELVSEAMTVLNKKNLTLSALRLFLQNVVVTNEVDLLTEEELEK 60

Query: 61 EKLFKQFQAEINKNIEDVRQGKFYTSEEVRSELGL 95
       EKLFKQFQAEINKNIEDVRQGKFYTSEEVR+ELGL
      Sbjct: 61 EKLFKQFQAEINKNIEDVRQGKFYTSEEVRAELGL 95
```

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 184

A DNA sequence (GBSx0190) was identified in *S.agalactiae* <SEQ ID 609> which encodes the amino acid sequence <SEQ ID 610>. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45      bacterial cytoplasm --- Certainty=0.4568(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

50 A related GBS nucleic acid sequence <SEQ ID 9513> which encodes amino acid sequence <SEQ ID 9514> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

-260-

>GP:CAA46375 GB:X65276 ORF A1 [Clostridium acetobutylicum]
Identities = 36/91 (39%), Positives = 51/91 (55%)

5 Query: 2 MSQIKLTPPEELRISAQKYTTGSQSITDVLTVI TQEQAVIDENWDGTA FDSFEAQFNELSP 61
M+QI +TPEEL+ AQ Y + I + + + I E W G AF ++ Q+N+L
Sbjct: 1 MAQISVTPEELKSQAQVYIQSKKEIDQAIQKVNSMNSTIAEEWKGQAFQAYLEQYNQLHQ 60

Query: 62 KITQFAQLLEDINQQLKVADVVEQTDSIA 92
+ QF LLE +NQQL K AD V + D+ A
10 Sbjct: 61 TVVQFENLLESVNQQLNKYADTV AERDAQDA 91

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 185

A DNA sequence (GBSx0191) was identified in *S.agalactiae* <SEQ ID 611> which encodes the amino acid sequence <SEQ ID 612>. Analysis of this protein sequence reveals the following:

Possible site: 21
>>> Seems to have no N-terminal signal sequence
20 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.4523 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 186

A DNA sequence (GBSx0192) was identified in *S.agalactiae* <SEQ ID 613> which encodes the amino acid sequence <SEQ ID 614>. Analysis of this protein sequence reveals the following:

Possible site: 44
>>> Seems to have no N-terminal signal sequence
35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.5339 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 187

A DNA sequence (GBSx0193) was identified in *S.agalactiae* <SEQ ID 615> which encodes the amino acid sequence <SEQ ID 616>. This protein is predicted to be chromosome assembly protein. Analysis of this protein sequence reveals the following:

5 Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.4620 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 188

A DNA sequence (GBSx0194) was identified in *S.agalactiae* <SEQ ID 617> which encodes the amino acid sequence <SEQ ID 618>. Analysis of this protein sequence reveals the following:

20 Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25 bacterial cytoplasm --- Certainty=0.4511 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 189

A DNA sequence (GBSx0195) was identified in *S.agalactiae* <SEQ ID 619> which encodes the amino acid sequence <SEQ ID 620>. Analysis of this protein sequence reveals the following:

35 Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
40 bacterial cytoplasm --- Certainty=0.5249 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 190

A DNA sequence (GBSx0196) was identified in *S.agalactiae* <SEQ ID 621> which encodes the amino acid sequence <SEQ ID 622>. Analysis of this protein sequence reveals the following:

Possible site: 14
 5 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3542 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 10 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9515> which encodes amino acid sequence <SEQ ID 9516> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 191

A DNA sequence (GBSx0197) was identified in *S.agalactiae* <SEQ ID 623> which encodes the amino acid sequence <SEQ ID 624>. Analysis of this protein sequence reveals the following:

Possible site: 15
 20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3098 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 192

A DNA sequence (GBSx0198) was identified in *S.agalactiae* <SEQ ID 625> which encodes the amino acid sequence <SEQ ID 626>. This protein is predicted to be rgg protein. Analysis of this protein sequence reveals the following:

Possible site: 59
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3177 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA26968 GB:M89776 rgg [Streptococcus gordonii]
 Identities = 74/277 (26%), Positives = 142/277 (50%)

-263-

Query: 7 IFREFRLNRQFSLKQVASNELSVSQLSRFERGESDLSLTKFLGALEAIDL SISEFMDRVN 66
 I + R ++ SLK+VA+ ++SV+QLSR+ERG S L++ F L + +S++EF +
 Sbjct: 10 ILKIIRESKNMSLKEVAAGDISVAQLSRYERGISSLTVDSFYSCRNMSVSLAEFQVVH 69

5 Query: 67 KYQKSDQISLMSQMAQHYQRDVAGLEKMSVEEGKLKKDSSDIRCRLNIVLFRGMICEC 126
 Y+++D + L +++++ + ++ LE +++ E ++ +LN ++ R + C
 Sbjct: 70 NYREADDVVLSQLSEAQRENNIVKLESILAGSEAMAQEFPEKKNYKLNTIVIRATLTSC 129

10 Query: 127 DSSRKMSSEEDLCFLSDYLQKDSWEISDYIILIGNLYRYYNTRHICQLVKEVINQKEYRD 186
 + ++S+ D+ FL+DYL F + W + L N + E+IN+ ++Y +
 Sbjct: 130 NPDYQVSKGDIEFLTDYLPSVEEWGRYELWLFTNSVNLLTLETFASEMINRTQFYNN 189

15 Query: 187 IYTNRNVVEATLLNVETLIERRALEEATFFLEKVEALLNNERNAYHRIILLYEKGLAY 246
 + NR + LLNVV IE L+ A FL ++ E + Y R+++ Y K +Y
 Sbjct: 190 LPENRRRIIKMLLNVVSACIENNHLQVAMKFLNYIDNTKIPETDLYDRVLIKYHKALYSY 249

20 Query: 247 AKGDSRGIQSMKQAIFCFQAIIGSKHHVENFQEHFNRV 283
 G+ ++Q + F+ + S +E F R+
 Sbjct: 250 KVGNPHARHDIEQCLSTFEYLDLDSFGVARKLKEQFERI 286

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 627> which encodes the amino acid sequence <SEQ ID 628>. Analysis of this protein sequence reveals the following:

25 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3792 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 79/275 (28%), Positives = 146/275 (52%), Gaps = 11/275 (4%)

35 Query: 9 REFRLNQFSLKQVASNELSVSQLSRFERGESDLSLTKFLGALEAIDL SISEFMDRVNKY 68
 R R +Q S+ +A LS SQ+SRFERGES+++ ++ L L+ +++++I EF+ +K
 Sbjct: 15 RRLRKKGKVSISSLADEYLSKSQISRFERGESEITCSRLLNLKDNLINITIDEFVSAHSKT 74

40 Query: 69 QKSDQISLMSQMAQHYQRDVAGLEKMSVEEGKLKKDSSDIRCRLNIVLFRGMICECD 128
 + +L+SQ + + +++V L K++ + KD R + +LF DS
 Sbjct: 75 H-THFFTLLSQARKCYAEKNNVVKLTKLL---KDYAHKDYE--RTMIKAILF----SIDS 123

45 Query: 129 SRKMSEEDLCFLSDYLQKDSWEISDYIILIGNLYRYYNTRHICQLVKEVINQKEYRDIY 188
 S S+E+L L+DYL F + W + IL+GN R+ N + L KE++ Y
 Sbjct: 124 SIAPSQEELTRLTDYLFKVEQWGYYEILLGNCSRMMNTLFLLTKEMVASFAYSEQNK 183

50 Query: 189 TNRNVVEATLLNVETLIERRALEEATFFLEKVEALLNNERNAYHRIILLYEKGLAYAK 248
 TN+ +V +N + I+ E + + + K++ LL +E N Y + + LY G+ +
 Sbjct: 184 TNKMLVTQLSINCLIIISIDHSCFEHSRYLINKIDLLLRLDELFNFYEKTVFLYVHGYYKLQ 243

Query: 249 GDSRGIQSMKQAIFCFQAIIGSKHHVENFQEHFNRV 283
 + G + M+QA+ F+ +G +++EH+ ++
 Sbjct: 244 EEMSGEEDMRQALQIFKYLGEDSLYYSYKEHYRQI 278

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 193

A DNA sequence (GBSx0199) was identified in *S.agalactiae* <SEQ ID 629> which encodes the amino acid sequence <SEQ ID 630>. This protein is predicted to be permease. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence

	INTEGRAL	Likelihood = -8.07	Transmembrane	217 - 233 (215 - 238)
	INTEGRAL	Likelihood = -7.96	Transmembrane	163 - 179 (158 - 185)
5	INTEGRAL	Likelihood = -7.75	Transmembrane	71 - 87 (69 - 91)
	INTEGRAL	Likelihood = -7.22	Transmembrane	369 - 385 (356 - 389)
	INTEGRAL	Likelihood = -5.15	Transmembrane	279 - 295 (275 - 299)
	INTEGRAL	Likelihood = -4.88	Transmembrane	252 - 268 (250 - 270)
10	INTEGRAL	Likelihood = -4.78	Transmembrane	140 - 156 (139 - 157)
	INTEGRAL	Likelihood = -3.56	Transmembrane	343 - 359 (340 - 367)
	INTEGRAL	Likelihood = -3.13	Transmembrane	40 - 56 (39 - 56)
	INTEGRAL	Likelihood = -2.28	Transmembrane	94 - 110 (92 - 112)

----- Final Results -----

15 bacterial membrane --- Certainty=0.4227(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAD36408 GB:AE001788 permease, putative [Thermotoga maritima]
 Identities = 97/396 (24%), Positives = 194/396 (48%), Gaps = 15/396 (3%)

Query: 1 MNINGIKLLSSRAVSKLGDVFYDYGNSWIASMGGQKILGIYQIVELLVSIVLNPFGG 60
 MN N + S VS +G Y + W+ S G + + G++ I L +I+++PF G

25 Sbjct: 1 MNRNLLLFAASGSFVSLIGTRIYQVALAWWLYSKTGSSEYV-GLFMISFLPAIIIVSPFAG 59

Query: 61 ALADRFQRRKILLITDAICAIM---CFLLSFIGDDKVMVYGLIVANAILAVSNAFSSPAY 117
 + DR RR ++++ D + ++ FL+ + + + L++ ++++V ++F +PA

Sbjct: 60 TVVDRHSRRNMMVVMDILRGVLFLMEYFSELTMAL--LLIVTVLVSVFDSFFNPBV 117

30 Query: 118 KSYIPEIVDKADIITYNANLETIVQIISVSSPVLGFLIFNNFGIRITLVIDAITFLISFL 177
 S +P++V K +++ N+ + + + P LG L+ G+ ++++++FLIS +
 Sbjct: 118 DSLLPDILVRKENLVRANSLYRLLKNLSKILGPALSLLKVVGLAGVILINSLSFLISGI 177

35 Query: 178 FLYAIKVERVQLSKQEKAIAKNILADIADGFTYIKKEKEIMFFLIIAALLNTFLAMFNYL 237
 F IKVE L K K +N+ DI YI+ + I+ +++ A++N F + L
 Sbjct: 178 FEMFIKVEEKHLKKVSK--RNMWQDIKSALLYRSVRFLVTILVIAIMNFFTGSMHVL 235

40 Query: 238 LP-FTNSLLKTSGAYATILSISAIGSIIGALIARKI--KSSINSMLSMLFSSLGVIVMG 294
 LP + L K+ Y T++S+ + G +I + I + +S+ ++ LV L V V
 Sbjct: 236 LPEHVSCLGKSEWWVYGTLMMSMLSFGGILIVTFILMATIRTRASVKTLGLNLVGYGLAVFVFA 295

45 Query: 295 FPSLFELPIWIPIYSGSFLFNSLLTMFNIHFPSQVQIRVDEAYMGRVMSTIFTIAIMFMPI 354
 W+ ++ FL T+FNI+ + +Q+ + E G++ S I ++ +P+
 Sbjct: 296 MTGNH---WLMFAMYFLIGIFQTLFNINVITLLQLAPEEMRGKIFSLSLISAVSFSLPV 351

50 Query: 355 GTLFMTIFSFSALSNVSFIVIGCAIAILGGLGFSYSK 390
 F S ++ + I GG+ S +
 Sbjct: 352 SYGFFGFLSSYVATAHIFITTSMALIAGGVILSLQR 387

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 631> which encodes the amino acid sequence <SEQ ID 632>. Analysis of this protein sequence reveals the following:

Possible site: 45

>>> Seems to have no N-terminal signal sequence

	INTEGRAL	Likelihood = -8.17	Transmembrane	172 - 188 (161 - 194)
	INTEGRAL	Likelihood = -8.07	Transmembrane	220 - 236 (218 - 242)
55	INTEGRAL	Likelihood = -7.22	Transmembrane	311 - 327 (303 - 329)
	INTEGRAL	Likelihood = -5.26	Transmembrane	98 - 114 (96 - 118)
	INTEGRAL	Likelihood = -4.99	Transmembrane	347 - 363 (342 - 370)
60	INTEGRAL	Likelihood = -4.62	Transmembrane	154 - 170 (151 - 171)
	INTEGRAL	Likelihood = -4.25	Transmembrane	284 - 300 (281 - 306)
	INTEGRAL	Likelihood = -3.66	Transmembrane	378 - 394 (378 - 396)
	INTEGRAL	Likelihood = -3.56	Transmembrane	74 - 90 (73 - 92)
	INTEGRAL	Likelihood = -2.39	Transmembrane	50 - 66 (49 - 66)

-265-

----- Final Results -----

bacterial membrane --- Certainty=0.4270 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

>GP:AAD36408 GB:AE001788 permease, putative [Thermotoga maritima]
 Identities = 85/345 (24%), Positives = 171/345 (48%), Gaps = 8/345 (2%)

10 Query: 40 SLSLVAVYQSLESVIGVLFNLFGGVIADSFRKKIIIITNILCGTACLVLSFLTKEQWLV 99
 S V ++ + ++ + F G + D R+ +++ +IL G + L + L
 Sbjct: 36 SSEYVGLPMISSFLPAIIIVSPFAGTVVDRHSRRNMMVVMDILRGVLFMYLFLMEYFSELT 95

15 Query: 100 YAIVL-TNVILAFMSAFSSPSYKAFTKEIVKKDSISQLNSLLETTSTVIKTVPMVAIFL 158
 A++L V+++ +F +P+ + ++V+K+++ + NSL + K+ P + L
 Sbjct: 96 MALLLIVTVLVSVFDSFFNPNAVDSDLPLVRKENLVRANSLYRLLKNLSKILGPALGSLL 155

20 Query: 159 YKLLGIGHGVLLLDGLSFLIIALLISFILEPVNDEVVIKEKVTIREFNDLKIGFKYVYSHK 218
 K++G+ GV+L++ LSFLI+ + FI +E +K+ R ++ D+K Y+ S +
 Sbjct: 156 LKVVGLAGVILINSLSFLISGIFEMFIKV--EIKHLKKVSKERNMWQDIKSALLYIRSVR 213

25 Query: 219 SIFIITVLSALVNFFLAAYNLLLPYSNQMFGEISTGLYGTFLTAAEAGFIGAILSGFVN 278
 I + ++ A++NFF + ++LLP G+ S +YGT ++ + GG I L +
 Sbjct: 214 FILVTILVIAIMNFFTGSMHVLLPEHVSKLGK-SEWVYGTMSMLSFGGLIVTFLMATIR 272

30 Query: 279 KELSSMRLLTFLSLSGLMLMLAPPFYIMFHNAIILALSPALFSLFLSIFNIQFFSLVQKD 338
 S L L GL + + + M N ++ L +F ++FNI +L+Q
 Sbjct: 273 TRASVKTTLGLNLVGYGLAVFV---FAMTGNHWLMFAMYFLIGIFQTLFNINVITLLQLA 328

25

Query: 339 VDNDFLGRVFGIIFTITILFMPIGTGFFSVALPNNNFSNFLFIIGS 383
 + + G++F +I ++ +P+ GFF + + ++FI S
 Sbjct: 329 IPEEMRGKIFSLISAVSFSLLPVSYGFGLSSYVATAHIFTTS 373

An alignment of the GAS and GBS proteins is shown below:

35 Identities = 136/379 (35%), Positives = 229/379 (59%), Gaps = 6/379 (1%)

Query: 8 LLSSRAVSKLGDVFYDYGNSTWIASMGGLGQKILGIYQIVELLVSIVLNPFGGALADRQ 67
 L+ S+ + ++GDV +D+ N+T++A + ++ +YQ +E ++ ++ N FGG +AD F+
 Sbjct: 11 LVYSKVIYRIGDVMFDFANNTFLAGLNPASLSLVAVYQSLESVIGVLFNLFGGVIADSKF 70

40 Query: 68 RRKILLITDAICAIMCFLLSFIGDDKVMVYGLIVANAILAVSNAFSSPAYKSYIPEIVDK 127
 R+KI++ T+ +C C +LSF+ ++ +VY ++ N ILA +AFSSP+YK++ EIV K
 Sbjct: 71 RKKIIITTNILCGTACLVLSFLTKEQWLVYAIVLTNVILAFMSAFSSPSYKAFTKEIVKK 130

45 Query: 128 ADIITYNANLETIVQIISVSSPVLGFLIFNNFGIRITLIVDAITFLISFLFLYAIKVERV 187
 I N+ LET +I V+ B++ + GI L+D ++FLI+ L + I
 Sbjct: 131 DSISQLNSLLETTSTVIKTVPMVAIFLYKLLGIHGVLVLLDGLSFLIAALLISFILPVND 190

50 Query: 188 QLSKQEKAIAKNILADIADGFTYIKKEKEIMFFLIIAALNLTFLAMFNLLPFTNSLLK- 246
 ++ +EKV I+ I D+ GF Y+ K I ++AL+N FLA +N LLP++N +
 Sbjct: 191 EVVIKEKVTTIREIFNDLKIGFKYVYSHKSIFIITVLSALVNFFLAAYNLLLPYSNQMFGE 250

55 Query: 247 -TSGAYATILSISAIGSIITGALIARKIKSSINSMLSMLVFSSILGVIVMGFPS---LFELP 302
 ++G Y T L+ AIG IGA++ + + +SM +L S G++M P +F
 Sbjct: 251 ISTGLYGTFLTAAEAGGGFIGAILSGFVNKEISSLRLILFLSLSGLMLMLAPPFYIMFHNA 310

60 Query: 303 IWIPYSGSFLFNSLLTMFNIHFFSQVQIRVDEAYMGRVMSTIFTIAIMFMPIGTLFMTIF 362
 I + S + LF+ I++FNI FFS VQ VD ++GRV IFTI I+FMPIGT F ++
 Sbjct: 311 IILALSPA-LFSLFLSIFNIQFFSLVQKDWDNDLGRVFGIIFTITILFMPIGTGFFSA 369

Query: 363 SFALSNVSFIVIGCAIAIL 381
 ++ + +IG I L
 Sbjct: 370 LNPNNSFNLFIIGSCITTL 388

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 194

A DNA sequence (GBSx0200) was identified in *S.agalactiae* <SEQ ID 633> which encodes the amino acid sequence <SEQ ID 634>. This protein is predicted to be membrane permease OpuCD. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL Likelihood = -5.68 Transmembrane 91 - 107 ( 88 - 110)
    INTEGRAL Likelihood = -4.30 Transmembrane 15 - 31 ( 9 - 37)
    INTEGRAL Likelihood = -3.72 Transmembrane 72 - 88 ( 72 - 88)
    INTEGRAL Likelihood = -3.19 Transmembrane 124 - 140 ( 123 - 142)

----- Final Results -----
bacterial membrane --- Certainty=0.3272(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8509> which encodes amino acid sequence <SEQ ID 8510> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1 Crend: 1
McG: Discrim Score: -10.69
GvH: Signal Score (-7.5): -3.79
Possible site: 39
>>> Seems to have no N-terminal signal sequence
ALOM program count: 5 value: -9.02 threshold: 0.0
    INTEGRAL Likelihood = -9.02 Transmembrane 35 - 51 ( 25 - 53)
    INTEGRAL Likelihood = -5.68 Transmembrane 151 - 167 ( 148 - 170)
    INTEGRAL Likelihood = -4.30 Transmembrane 75 - 91 ( 69 - 97)
    INTEGRAL Likelihood = -3.72 Transmembrane 132 - 148 ( 132 - 148)
    INTEGRAL Likelihood = -3.19 Transmembrane 184 - 200 ( 183 - 202)
    PERIPHERAL Likelihood = 2.17      58
modified ALOM score: 2.30

----- Final Results -----
bacterial membrane --- Certainty=0.4609(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF91342 GB:AF249729 membrane permease OpuCD [Listeria monocytogenes]
Identities = 104/154 (67%), Positives = 133/154 (85%)
Query: 3 IANVIQTIPSLAMISIIMLGLGLIKTVVATVFLYSLLPIITNTYTGIRNVDSDLLDAAK 62
        IAN+IQTIP+LAM+++ML +GLG TVV ++FLYSLLPI+ NTYTGIRNVD LL++ K
Sbjct: 60 IANIIQTI PALAMI LAVMLI MGLGTNTVVLSLFLY SLLPILKNTYTGIRNVDGALLES GK 119

Query: 63 GMGMTKQRQLFMVELPLSISVIMAGL RNALVVAI GITAI GAFVGGGLGDIIIRGTNATN 122
        MGMTK Q L ++E+PL++ SVIMAG+RNALV+AIG+ AIG FVG GGLGDII+RGTNATN
Sbjct: 120 AMGMTK WQVRLTEMPLALS VIMAGI RNALVIAIGVAAIGTFVGAGGLDII VRGTNATN 179

Query: 123 GGAIILLAGSLPTALMAIFSDLILGGIQRML EPRK 156
        G AIILAG++PTA+MAI +D++LG ++R L P K
Sbjct: 180 GTAIILAGAIPTAVMAILADVLLGWVERTLNPK 213
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 635> which encodes the amino acid sequence <SEQ ID 636>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

5	INTEGRAL Likelihood = -9.24 Transmembrane 39 - 55 (31 - 59)
	INTEGRAL Likelihood = -7.17 Transmembrane 190 - 206 (188 - 211)
	INTEGRAL Likelihood = -4.62 Transmembrane 93 - 109 (75 - 110)
	INTEGRAL Likelihood = -3.66 Transmembrane 76 - 92 (75 - 92)
	INTEGRAL Likelihood = -2.87 Transmembrane 221 - 237 (220 - 237)
	INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 (165 - 184)

10 ----- Final Results -----
 bacterial membrane --- Certainty=0.4694 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

15 The protein has homology with the following sequences in the databases:

>GP:AAD45530 GB:AF162656 choline transporter [Streptococcus pneumoniae]
 Identities = 344/508 (67%), Positives = 425/508 (82%), Gaps = 2/508 (0%)

20	Query: 13 MPSLFVTFQNRFNEWLAAALGEHLQISLLSILMIALLIGVPLAALLSRSKRWSDIMLQVTGV 72 M +L TFQ+RF++WL AL +HLQ+SLL+L++A+L+ +PLA L ++ +D +LQ+ G+
	Sbjct: 1 MTNLIATFQDRFSDWLTALSQHQLQLSLLTLLAAILAIPLAVFLRYHEKLADWVLQIAGI 60
25	Query: 73 FQTIPSLALLGLFIPLMGIGTLPAPVTALVIYIAIFPILQNTITGLNGIDPSLVEAGIAFGM 132 FQTIPSLALLGLFIPLMGIGTLPAP+TALVIYIAIFPILQNTITGL GIDP+L EAGIAFGM Sbjct: 61 FQTIPSLALLGLFIPLMGIGTLPALTALVIYIAIFPILQNTITGLKGIDPNLQEAGIAFGM 120
30	Query: 133 TKWERLKTFEIPIAMPVIMSGVRTSAVMIIIGTATLASLIGAGGLGSFILLGIDRNNANLI 192 T+WERLK FEIP+AMPVIMMSG+RT+AV+IIGTATLA+LIGAGGLGSFILLGIDRNNNA+LI Sbjct: 121 TRWERLKKFEIPLAMPVIMSGIRTAAVLIIGTATLAALIGAGGLGSFILLGIDRNNASLI 180
35	Query: 193 LIGAISALLAIIFNSLLQYLEKASLRRIMISFGITLLALLASYTPMALSQFSKGKDITVV 252 LIGA+SSA+LAI FN LL+ +EKA LR I F + L L SY+P L Q K K+ +V Sbjct: 181 LIGALSSAVLATAFNFLVKMEAKLRTIFSGFALVALLGLSYSPALLVQ--KEKENLV 238
40	Query: 253 IAGKLGAEPDILINLYKELIEDQSDISVELKSNSFGKTSFLYEALKSGDIDMYPEFTGTIT 312 IAGK+G EP+IL N+YK LIE+ + ++ +K NFGKTSFLYEALK GDID+YPEFTGT+T Sbjct: 239 IAGKIGPEPEILANMYKLLIEENTSMATVKPNFGKTSFLYEALKKGDIDIYYPEFTGTITV 298
45	Query: 313 SSLLRDKPPLSNDPKQVVEDAKKGIAKQDKLTLLKPFAYQNTYAVAMPEKLAKEYQIETI 372 SLI+ P +S++P+QVY+ A+ GIAKD L LKP +YQNTYAVA+P+K+A+EY ++TI Sbjct: 299 ESLQPSPKVSHEPEQVYQVARDGIAKQDHAYLKPMQSYQNTYAVAVPKKIAQEYGLKTI 358
50	Query: 373 SDLKAHADTLKAGFTLEFKDRADGYKGMQSQYGLQLSVATMEPALRYQAIQSGDIQVTD 432 SDLK LKAGFTLEF DR DG KG+QS YGL L+VAT+EPALRYQAIQSGDIQ+TDA Sbjct: 359 SDLKKVEGQLKAGFTLEFNDREDGNKGLQSMYGLNLNVATIEPALRYQAIQSGDIQITDA 418
55	Query: 433 YSTDAAEITKYHLKVLKDDKQLFPFYQGAPLMKTSLLTKHPELKGINQLAGKITEKEMQD 492 YSTDAAE+ +Y L+VL+DDKQLFPFYQGAPLMK +LL KHPPEL+ +LN LAGKITE +M Sbjct: 419 YSTDAAEELERYDLQVLEDDKQLFPFYQGAPLMKEALLKKHPELERVLNTLAGKITESQMSQ 478
	Query: 493 MNYEVSVKGADANKVARDYLLKTGLIQQ 520 +NY+V V+G A +VA+++L + GL++K Sbjct: 479 LNYQVGVEGKSAKQVAKEFLQEQQGLKK 506

55 An alignment of the GAS and GBS proteins is shown below:

Identities = 53/148 (35%), Positives = 93/148 (62%), Gaps = 1/148 (0%)

60	Query: 3 IANVIOTIPSLAMISIIMLGLGLGIKTVVATVFLYSLLPIITNTYTGIRNVDSDLLDAAK 62 + V QTIPSLA++ + + +G+G V + +Y++ PI+ NT TG+ +D L++A Sbjct: 69 VTGVFQTIPSLALLGLFIPLMGIGTLPAPVTALVIYIAIFPILQNTITGLNGIDPSLVEAGI 128
65	Query: 63 GMGMTKQRQLFMVELPLSISVIMAGLRLNALVVAIGITAIGAFVGGGGDIIIRGTNATN 122 GMTK +RL E+P+++ VIM+G+R + V+ IG + + +G GGLG I+ G + N Sbjct: 129 AFGMTKWERLKTFEIPIAMPVIMSGVRTSAVMIIIGTATLASLIGAGGLGSFILLGIDRNN 188
	Query: 123 GGAIILAGSLPTALMAIFSDLILGGIQR 150

+IL G++ +AL+AI + +L +++
 Sbjct: 189 AN-LILIGAIS ALLAI IFNS LLQYLEK 215

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 5 vaccines or diagnostics.

Example 195

A DNA sequence (GBSx0201) was identified in *S.agalactiae* <SEQ ID 637> which encodes the amino acid sequence <SEQ ID 638>. This protein is predicted to be choline transporter-related. Analysis of this protein sequence reveals the following:

10 Possible site: 44
 >>> May be a lipoprotein
 INTEGRAL Likelihood = -3.03 Transmembrane 306 - 322 (306 - 327)
 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.2211(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9517> which encodes amino acid sequence <SEQ ID 9518>
 20 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB15386 GB:Z99121 glycine betaine/carnitine/choline ABC
 transporter (osmoprotectant-binding protein) [Bacillus subtilis]
 Identities = 168/303 (55%), Positives = 224/303 (73%), Gaps = 1/303 (0%)
 25 Query: 2 LKKSHFLQIIFTLCLALLTISGCGQLTDKKSHTTIKVAQSSTESSIMANIITELIHHEL 61
 + K +L F L +L + GC L + TIK+ AQS TES I+AN+I +LI H+
 Sbjct: 1 MTKIKWLGAFA LVFVML-LGGCSLPGLGGASDDTIKIGAQSMTESEIVANMIAQLIEHDT 59
 30 Query: 62 GYNNTLISNLGSSTVTHQALLRGDADIAATTRYTGTDITGTLGLKAVKDPKEASKIVKTEF 121
 NT L+ NLGS+ V HQA+L GD DI+ATRY+GTD+T TLG +A KDPK+A IV+ EF
 Sbjct: 60 DLNTALVKNLGSNYVHQAM LGGDIDISATRYSGTDLTSLGKEAEKDPKKALNIVQNEF 119
 35 Query: 122 QKRYNQTWYPTYGFSDTYAFMVTKEFARQNKITKISDLKKLSTTMKAGVDSSWMNRREGDG 181
 QKR++ W+ +YGF +TYAF VTK+FA + I +SDLKK ++ K GVD++W+ R+GDG
 Sbjct: 120 QKRF SYKWFDSYGF DNTYAF TVKKFAEKEHINTVSDLKKNASQYKLGV DNAWLKRKGDG 179
 40 Query: 182 YTDFAKTYGF EF SHI YPMQ IGLVYDAVE SNKM QSVLG YSTDGR ISYYDLE ILR DDK KFFP 241
 Y F TYG FEF YPMQ IGLVY DA++ KM +VL YSTDGRI +YDL+IL+DDK+FFF
 Sbjct: 180 YKGF VSTY GF EFG TTYP MQI GLVY DAVK NGKMDA VLAY STDGRI KAYDL KILK DDKR FFP 239
 45 Query: 242 PYEASMVVNN SIIKKDPKL KKLLHRLDGKINLKT M QNL N YM VDDKL LEPSVVA KQF LEKN 301
 PY+ S V+ ++K+ P+L+ +++++L G+I+ +TMQ L NY VD KL EPSVVA K+F LEK+
 Sbjct: 240 PYDCSPVIPEKVLKEHPELEGVINKLIGQIDTETM QEL NYEV DGKL KEPSVVA K FLEKH 299
 Query: 302 HYF 304
 HYF
 Sbjct: 300 HYF 302

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8511> and protein <SEQ ID 8512> were also identified. Analysis of this protein sequence reveals the following:

55 Lipop: Possible site: 22 Crend: 5
 McG: Discrim Score: 10.26
 GvH: Signal Score (-7.5): -4.19

-269-

Possible site: 44
 >> May be a lipoprotein
 ALOM program count: 0 value: 8.65 threshold: 0.0
 PERIPHERAL Likelihood = 8.65 66
 5 modified ALOM score: -2.23

----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 10 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

56.3/75.4% over 287aa

Bacillus subtilis

15 EGAD|109208| glycine betaine/carnitine/choline ABC Insert characterized
 SP|O32243|OPCC_BACSU GLYCINE BETAINE/CARNITINE/CHOLINE-BINDING PROTEIN PRECURSOR
 (OSMOPROTECTANT-BINDING PROTEIN). Insert characterized
 GP|2635894|emb|CAB15386.1||Z99121 glycine betaine/carnitine/choline ABC transporter
 20 (osmoprotectant-binding protein) Insert characterized
 PIR|E69670|E69670 glycine betaine/carnitine/choline ABC transporter (osmoprotec) opuCC -
 Insert characterized

25 ORF01181(349 - 1212 of 1524)
 EGAD|109208|BS3376(15 - 302 of 303) glycine betaine/carnitine/choline ABC {Bacillus subtilis} SP|O32243|OPCC_BACSU GLYCINE BETAINE/CARNITINE/CHOLINE-BINDING PROTEIN PRECURSOR (OSMOPROTECTANT-BINDING PROTEIN). GP|2635894|emb|CAB15386.1||Z99121 glycine betaine/carnitine/choline ABC transporter (osmoprotectant-binding protein) {Bacillus subtilis} PIR|E69670|E69670 glycine betaine/carnitine/choline ABC transporter (osmoprotec) opuCC - Bacillus subtilis
 %Match = 33.5
 %Identity = 56.2 %Similarity = 75.3
 Matches = 162 Mismatches = 71 Conservative Sub.s = 55

35 162 192 222 252 282 312 342 372
 VVVFFLIVF*QCLIFIFSVRYKSGSMKRIWGVXXN*LXXITGNSSNAQNNKKGGLDMLKKSHFLQIFTLCALLTISGCQ
 : : : : |||
 MTKIKWLGAFAVLFVVMLLGGCS
 10 20

40 402 432 462 492 522 552 582 612
 LTDTKKGHTTIKVAAQSSTESSIMANIITELIHHLEGNTTLISNLGSSTVTHQALLRGDADIAATRTYGTIDITGTLGL
 | : |||: ||| ||| ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
 LPGLGGASDDTIKIGAQSMTESEIVANMIAQLIEHDTDLNTALVKNLGSNYVQHQAMLLGGDIDISATRYSQTDLTSTLGK
 45 40 50 60 70 80 90 100

50 642 672 702 732 762 792 822 852
 KAVKDPEASKIVKTEXQKRYNQTWYPTYGFSDTYAFMVTKEARQNKITKISDLKKLSTMKAGVDSSWMNREGDGYTD
 :| ||||:| |||:| |||:| :| :|||:| |||:| |||:| :| :|||:| :| :|||:| :| :|||:
 EAEKDPKKALNIVQNEFQKRFSYKWFDSYGFDNITYAFTVTKFAEKEHINTVSDLKKNASQYKLGVDNAWLKRKGDGYKG
 120 130 140 150 160 170 180

55 882 912 942 972 1002 1032 1062 1092
 FAKTYGFEGFSHIYPMQIGLVYDAVESENKMQSVLGYSTDGRISYYDLEILRDKKFPPYEASMVNNNSIIKKDPKLKLL
 | |||||:| |||:| |||:| |||:| :| |||:| :| |||:| :| |||:| :| |||:| :| :|||:
 FVSTYGFEFGTTYPMQIGLVYDAVKNGKMDAVLAYSTDGRIKAYDLKILKDDKRFPPYDCSPVIPEKVLKEHPELEGVI
 200 210 220 230 240 250 260

60 1122 1152 1182 1212 1242 1272 1302 1332
 HRLDGKINLKTMQNLNYMVDDKLLEPSVVAQKFLEKNHYFRGDK*MKQMNTFQQFIYYFQHNGSYILEQFIIHHFLISVG
 ::| :||:| :||| ||| |||:| :|||:| :|||:
 NKLIGQIDTETMQUELYEVGDGKLKEPSVVAKEFLEKHHYFD
 280 290 300

65 SEQ ID 8512 (GBS23) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 8; MW 35kDa).

The GBS23-His fusion product was purified (Figure 194, lane 9) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 251). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Example 196

- 5 A DNA sequence (GBSx0202) was identified in *S.agalactiae* <SEQ ID 639> which encodes the amino acid sequence <SEQ ID 640>. This protein is predicted to be membrane permease OpuCB (opuBB). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
10    INTEGRAL    Likelihood = -9.66    Transmembrane  25 - 41 ( 18 - 45)
    INTEGRAL    Likelihood = -7.96    Transmembrane  182 - 198 ( 174 - 202)
    INTEGRAL    Likelihood = -4.83    Transmembrane  61 - 77 ( 57 - 95)
    INTEGRAL    Likelihood = -4.09    Transmembrane  78 - 94 ( 78 - 95)
    INTEGRAL    Likelihood = -1.22    Transmembrane  134 - 150 ( 134 - 150)

15      -----
      ----- Final Results -----
          bacterial membrane --- Certainty=0.4864 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

20 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF91340 GB:AF249729 membrane permease OpuCB [Listeria
monocytogenes]
Identities = 121/208 (58%), Positives = 160/208 (76%)
25
Query: 1 MVNFLSQYGMQILVKTWEQVYISFFAIALGIAIAVPLGVVLTRFPKVAKIIIASIQLQT 60
+V F + G +LV+TW+ ++IS A+ LGIA+AVP G++LTR PKVA +I + S+LQT
Sbjct: 4 IVTFQENGHNLLVQTWQHFLFISLSAVILGIAAVAVPTGILLTRSPKVANFVIGVVSVLQT 63

30
Query: 61 IPSLALLALMIPILFGIGKIPAIVALFIYSSLPILRNTYIGMNNVNPTLKDCAKGMGMKPI 120
+PSLA+LA +IP G+G +PAI+ALFIY+LLPILRNT+IG+ V+ L + +GMGM
Sbjct: 64 VPSLAILAFIIPFLGVGTLPATIALFIYALLPILRNTFIGVRGVVDKNLIESGRGMGMTNW 123

35
Query: 121 QSIFQVELPLATPIIMAGIRLSTIYVIAWATLASYIGAGGLGDLIFSGLNLFQSKLILGG 180
Q I VE+P + +IMAGIRLS +YVIWATLASYIGAGGLGD IF+GLNL++ LILGG
Sbjct: 124 QLIVNVEIPNSISVIMAGIRLSAVYVIAWATLASYIGAGGLGDFIFNGLNLYRPDLILGG 183

40
Query: 181 TIPVIIILSLIIDYLLGLLETALTPTTR 208
IPV IL+L++++ LG LE LTP+ R
Sbjct: 184 AIPVTILALVVEFALGKLEYRLTPKAIR 211
```

A related GBS gene <SEQ ID 8513> and protein <SEQ ID 8514> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1 Crend: 0
45  McG: Discrim Score: -9.08
  GvH: Signal Score (-7.5): -1.86
      Possible site: 37
>>> Seems to have no N-terminal signal sequence
ALOM program count: 5 value: -8.60 threshold: 0.0
50    INTEGRAL    Likelihood = -8.60    Transmembrane  25 - 41 ( 18 - 45)
    INTEGRAL    Likelihood = -7.96    Transmembrane  182 - 198 ( 174 - 202)
    INTEGRAL    Likelihood = -4.83    Transmembrane  61 - 77 ( 57 - 95)
    INTEGRAL    Likelihood = -4.09    Transmembrane  78 - 94 ( 78 - 95)
    INTEGRAL    Likelihood = -1.22    Transmembrane  134 - 150 ( 134 - 150)

55    PERIPHERAL Likelihood = 2.70      156
modified ALOM score: 2.22

*** Reasoning Step: 3
```

-271-

----- Final Results -----

bacterial membrane --- Certainty=0.4439 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the databases:

ORF01825 (301 - 927 of 1233)

GP|9651976|gb|AAF91340.1|AF249729_2|AF249729 (4 - 212 of 218) membrane permease OpuCB
 {Listeria monocytogenes}

10

%Match = 30.2

%Identity = 57.9 %Similarity = 79.9

Matches = 121 Mismatches = 42 Conservative Sub.s = 46

117 147 177 207 237 267 297 327

15

STCF*YLKTY*FLCYGRRLT*KYC*AYFKTIWFKIRSSC*P*E*LKGHCYSCIPS*YVIRYYLGRY*NGGSIMVNFLSQYG
 : | : : |
 MDAIVTFFQENG
 10

20

357 387 417 447 477 507 537 567

MQILVKTWEQVYISFFAIALGIAIAVXPXGVVLTRFPKVAKIIIAIASMLQTIPSLALLALMIPFGIGKIPAIVALFIYS
 : ||:||:::||: |: ||||:||| :||| : |:|||:|||:|||:|||:|||:
 HNLLVQTWQHLFISLSAVILGIAAVAVPTGILLTRSPKVANFVIGVVSVLQTVPSLAILAFIIPFLGVGTLPAIIALFIYA
 30 40 50 60 70 80 90

25

597 627 657 687 717 747 777 807

LLPILRNTYIGMNNVNPTLKDCAKGGMKPIQSFQVELPLATPIIMAGIRLSTIYVIAWATLASYIGAGGLGDLIFSGL
 |||||||:||: |: | : :||| | | |||: | : :|||||:|||:|||:|||:
 LLPILRNTFIGVRGVDKNLIESGRGMGMTNWQLIVNEIPNSISVIMAGIRLSAVYVIAWATLASYIGAGGLGDFIFNGL
 30 110 120 130 140 150 160 170

30

837 867 897 927 957 987 1017 1047

NLFQSKLILGGTIPVIILSLLIODYLLGLLETALTPRTTRREA*ICLNRTFYRLHFA*PS*RFLVVN*PILKSLVIPQL
 ||:: ||||| ||| ||:|::: ||| ||| |||:
 NLYRPDLILGGAIPVITILALVVEFALGKLEYRLTPKAIREAREGGE
 190 200 210

35

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 197

40

A DNA sequence (GBSx0203) was identified in *S.agalactiae* <SEQ ID 641> which encodes the amino acid sequence <SEQ ID 642>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

45

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3531 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF91339 GB:AF249729 ATPase OpuCA [Listeria monocytogenes]

Identities = 230/380 (60%), Positives = 298/380 (77%), Gaps = 4/380 (1%)

55

Query: 6 IIEYQNINKVY-GENVAVEDINLK1YPGDFVCFIGTSGSGKTTLMRMVNHMLKPTNGTLL 64
 ++++++ K Y G AV D+ L I G+FVFCFIG SG GKTT M+M+N +++PT G +

Sbjct: 1 MLKFEHVTKTYKGGKKAVNDLTLNIDKGEFVFCFIGPSGCGKTTTMKMINRLIEPTEGKIF 60

Query: 65 FKGKDISTINPIELRRIGYV1QNIIGLMPHMTIYEN1VLPKLLKWSEAKAKARELIK 124
 KDI +P++LRR IGYVIQ IGLMPHMTI EN1VLPKLLKWSEE K+ +A+ELIK

60

Sbjct: 61 INDKDIMAEDPVKLRRSIGYV1QQIGLMPHMTIREN1VLPKLLKWSEEKKQERAKELIK 120

-272-

Query: 125 LVELPEEYLDRYPTSELSGGQQQRIGVIRALAADQDIILMDEPPGALDPITREGIQDLVKS 184
 LV+LPEE+LDRYPT ELSGGQQQRIGV+RALAA+Q++ILMDEPPGALDPITR+ +Q+ K+
 Sbjct: 121 LVSDLPEEFLDRPYELSGGQQQRIGVRLAEEQNLLILMDEPPGALDPITRDSLQEEFKN 180

5 Query: 185 LQEEMGKTIILVTHDMDEALKLATKIIIVMDNGKMVQEGTPNDLLHHHPATSFVEQMIGER 244
 LQ+E+GKTII VTHDMDEA+KLA +I++M +G++VQ TP++L +PA SFVE IG++R
 Sbjct: 181 LQKELGKTIIFVTMDEAIKLADRIVIMKDGEIVQFDTPDEILRNPANSFVEDFIGKDR 240

10 Query: 245 LLHAQADITPVKQIMLNPNVSITAETKLTEAITLMRQKRVDSSLVTNDGKLI-GFIDLES 303
 L+ A+ D+T V QIM NPVSITA+K+L AIT+M++KRVD+LLV D G ++ GFID+E
 Sbjct: 241 LIEAKPDVTQVAQIMNTNPVSITADKSLQAAITVMKEKRVDTLLVVDEGNVLKGFDVEQ 300

15 Query: 304 LSSKYKKDRLVSDILKHDFYVMEDDLRLNATAERILKLGKLYAPVVDHENNLKGIVTRAS 363
 + + V DI++ FYV ED LLR+T +RILK G KY PVVD + L GIVTRAS
 Sbjct: 301 IDLNRRATSVMDIIEKNVFYVYEDTLRDTVQRILKRGYKYIPVVDKDKRLVGVTRAS 360

20 Query: 364 LVDMLYDIIWGDTE--TEDQ 381
 LVD++YD IWG E TE+Q
 Sbjct: 361 LVDIVYDSIWGTLIEDATENQ 380

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 643> which encodes the amino acid sequence <SEQ ID 644>. Analysis of this protein sequence reveals the following:

Possible site: 39
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3619 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below:

Identities = 102/237 (43%), Positives = 165/237 (69%), Gaps = 1/237 (0%)

35 Query: 6 IIYEQNINKVYGENVAVEDINLK IYPGDFVCFIGTSGSGKTTLMRMVNHMLKPTNGTLLF 65
 +I + N++K +G+ +++ +I +F +G SGSGKTTL++M+N +++P++G +L
 Sbjct: 1 MIRFNNVSKTFQTKVLQEQTQINDREFFVLVGPSGSKGTTLLKMINCLIEPSSGDILL 60

40 Query: 66 KGKDISTINPIELRRRIGYVIQNIQLMPHMITYENIVLVPKLLKWSEEAKRAKARELIK 125
 + ++ E+R IGYV+Q I L P++T+ ENI ++P++ +WS E R K EL+
 Sbjct: 61 NNVPQTELDLREMRLSISGYVLQQIALFPNLTVANIAIIPEMKQWSAEIRQKTEELDK 120

45 Query: 126 VELP-EEYLDRYPTSELSGGQQQRIGVIRALAADQDIILMDEPPGALDPITREGIQDLVKS 184
 V LP ++YLDRYPT+LSGG+QQRIG++RA+ + I+LMDEPF ALDPI+R+ +Q+L+ S
 Sbjct: 121 VGLPAKDYLDRYPTSDLSGGEQQRIGIVRAIISHPKILLMDEPFSALDPISRKQLQELMLS 180

Query: 185 LQEEMGKTIILVTHDMDEALKLATKIIIVMDNGKMVQEGTPNDLLHHHPATSFVEQMIG 244
 L +E TI+ VTHDMDEA+KLA +I++M +G++VQ P + HPA +FV + G
 Sbjct: 181 LHKEFDMTIVFTMDEAIKLGDRAVAILNEGEIVQLDRPEMIKTHPANAFVVLFG 237

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 198

A repeated DNA sequence (GBSx0212) was identified in *S.agalactiae* <SEQ ID 645> which encodes the amino acid sequence <SEQ ID 646>. Analysis of this protein sequence reveals the following:

55 Possible site: 24
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4736 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-273-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 199

A DNA sequence (GBSx0213) was identified in *S.agalactiae* <SEQ ID 647> which encodes the amino acid sequence <SEQ ID 648>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 38
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL Likelihood = -1.06 Transmembrane 18 - 34 ( 18 - 34)

15 ----- Final Results -----
    bacterial membrane --- Certainty=0.1426 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 A related GBS gene <SEQ ID 8515> and protein <SEQ ID 8516> were also identified. Analysis of this protein sequence reveals the following:

```

    Lipop: Possible site: 20 Crend: 5
    Sequence Pattern: CQMN
    SRCFLG: 0
25    McG: Length of UR: 19
    Peak Value of UR: 2.60
    Net Charge of CR: 3
    McG: Discrim Score: 7.77
    GvH: Signal Score (-7.5): -4.89
30    Possible site: 25
    >>> May be a lipoprotein
    Amino Acid Composition: calculated from 21
    ALOM program count: 0 value: 13.21 threshold: 0.0
    PERIPHERAL Likelihood = 13.21 115
35    modified ALOM score: -3.14

    *** Reasoning Step: 3

    ----- Final Results -----
40    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

45    ORF01527(346 - 465 of 1095)
    EGAD|7398|7198(2 - 41 of 47) lysis protein for colicin e9 precursor {Escherichia coli}
    EGAD|41475|43808 lysis protein { } SP|P13344|LYS5_ECOLI LYSIS PROTEIN FOR COLICIN E5
    PRECURSOR. GP|40543|emb|CAA33861.1||X15857 lysis protein (AA 1-47) {Enterobacteriaceae}
    GP|144373|gb|AAA98053.1||M30445 colicin release protein {Plasmid Cole5-099}
    PIR|JQ0330|JQ0330 colicin E5 lysis protein precursor - Escherichia coli plasmid Cole5-099
    %Match = 3.7
    %Identity = 35.0 %Similarity = 52.5
    Matches = 14 Mismatches = 19 Conservative Sub.s = 7

55    135      165      195      225      255      285      315      345
    YIYFFHCRRIYIIININY*FN*GI*NIQMIFCLHVKTKTIKIRENFVILKLIL*CW*IIVNFIYLIYKIVILRKENMMR

```

M

375 405 435 465 495 525 555 585
 KYIKWLIPISIFGMLGGCQMNSEHKIQSNEVKNSKQSEVKKDKKMTKEQLAYLKEHEQEIIDYVKLHNNQIESVQFDW
 | | | : : : || | | | | : | | | | | : |:
 KKJITWIIILLLAATIILAACQANYIHDVQGGTVSPSSSAELTGLATO
 20 30 40

- 10 SEQ ID 8516 (GBS389) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 6; MW 18kDa).

The GBS389-His fusion product was purified (Figure 214, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 313), which confirmed that the protein is immunoaccessible on GBS bacteria.

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 200

A DNA sequence (GBSx0214) was identified in *S.agalactiae* <SEQ ID 649> which encodes the amino acid sequence <SEQ ID 650>. Analysis of this protein sequence reveals the following:

20 Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25 bacterial cytoplasm --- Certainty=0.3766 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S. pyogenes*.

- 30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 201

A DNA sequence (GBSx0215) was identified in *S.agalactiae* <SEQ ID 651> which encodes the amino acid sequence <SEQ ID 652>. Analysis of this protein sequence reveals the following:

```
35      Possible site: 46
         >>> Seems to have no N-terminal signal sequence

         ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.3882 (Affirmative) < succ>
         bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
         bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 202

A DNA sequence (GBSx0216) was identified in *S.agalactiae* <SEQ ID 653> which encodes the amino acid sequence <SEQ ID 654>. This protein is predicted to be lectin, alpha subunit precursor. Analysis of this protein sequence reveals the following:

```

5    Possible site: 47
     >>> Seems to have no N-terminal signal sequence

     ----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.0653 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

15 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 203

A DNA sequence (GBSx0217) was identified in *S.agalactiae* <SEQ ID 655> which encodes the amino acid sequence <SEQ ID 656>. Analysis of this protein sequence reveals the following:

```

20    Possible site: 41
     >>> Seems to have no N-terminal signal sequence

     ----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.6569 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 204

A DNA sequence (GBSx0218) was identified in *S.agalactiae* <SEQ ID 657> which encodes the amino acid sequence <SEQ ID 658>. Analysis of this protein sequence reveals the following:

```

35    Possible site: 27
     >>> Seems to have no N-terminal signal sequence

     ----- Final Results -----
40      bacterial cytoplasm --- Certainty=0.5736 (Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-276-

Example 205

A DNA sequence (GBSx0219) was identified in *S.agalactiae* <SEQ ID 659> which encodes the amino acid sequence <SEQ ID 660>. Analysis of this protein sequence reveals the following:

Possible site: 52
 5 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -13.11 Transmembrane 146 - 162 (138 - 170)
 INTEGRAL Likelihood = -12.90 Transmembrane 13 - 29 (9 - 32)
 INTEGRAL Likelihood = -9.50 Transmembrane 108 - 124 (104 - 129)
 INTEGRAL Likelihood = -7.75 Transmembrane 40 - 56 (33 - 61)
 10 INTEGRAL Likelihood = -6.64 Transmembrane 177 - 193 (170 - 195)
 INTEGRAL Likelihood = -3.35 Transmembrane 77 - 93 (77 - 97)

 ----- Final Results -----
 bacterial membrane --- Certainty=0.6243 (Affirmative) < succ>
 15 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8517> which encodes amino acid sequence <SEQ ID 8518> was also identified.

20 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 206

25 A DNA sequence (GBSx0220) was identified in *S.agalactiae* <SEQ ID 661> which encodes the amino acid sequence <SEQ ID 662>. Analysis of this protein sequence reveals the following:

Possible site: 37
 >>> Seems to have no N-terminal signal sequence

 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2374 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB89623 GB:AE000990 repressor protein [Archaeoglobus fulgidus]
 Identities = 34/62 (54%), Positives = 46/62 (73%)

 40 Query: 11 LKQVREDIGMTQQELAIRIGVRRETIGHLENNRYNPSLEMALKIVKIFDMKIEDIFQLRK 70
 +K+ R MTQ+ELA R+GVRRETI LE +YNPSL++A KI ++F+ KIEDIF +
 Sbjct: 5 IKEFRAKFNMQTQEEELAKRKGVRRETIVFLEKGKYNPSLKLAYKIAVFNNAKIEDIFIFDE 64

 Query: 71 ED 72
 45 E+
 Sbjct: 65 EE 66

There is also homology to SEQ ID 412.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-277-

Example 207

A DNA sequence (GBSx0221) was identified in *S.agalactiae* <SEQ ID 663> which encodes the amino acid sequence <SEQ ID 664>. Analysis of this protein sequence reveals the following:

5 Possible site: 36

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3794 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

10 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB61817 GB:AL133236 putative acetyl transferase [Streptomyces
coelicolor A3(2)]

15 Identities = 30/97 (30%), Positives = 52/97 (52%), Gaps = 1/97 (1%)

Query: 82 VGMNLNIVTLARADMQWGEELGYVFHNQFWNSNGYAFESILALLNSTYEKLGFHHIEAQITPG 141
VGM ++ + Q GE+ Y+ H + W G E +LL+ +++ G H I A P

20 Sbjct: 72 VGMGDLHVRSHQRQ-GEISYIVHPRVWGQQGIGTEIGRSLLSLGFDRWGLHRIRATCDPR 130

Query: 142 NERSEKLVRRRLGLTYETTRKDFSFENGKWTDKLIYSI 178

N+ S +++ +LG+TYE + ++ W D L++SI

Sbjct: 131 NQASSRVLTKLGMTYEGRHRHTAWIRDGWRDSLVFSI 167

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 208

A DNA sequence (GBSx0222) was identified in *S.agalactiae* <SEQ ID 665> which encodes the amino acid sequence <SEQ ID 666>. This protein is predicted to be p20 protein. Analysis of this protein sequence reveals the following:

35 Possible site: 44

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1044 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA30415 GB:X07542 P20 (AA 1-178) [Bacillus licheniformis]

Identities = 56/175 (32%), Positives = 94/175 (53%), Gaps = 6/175 (3%)

45 Query: 16 TVLTERLRLQPVELTNVNDLFLEFSSDSETVFYMQRYKANTVEEAQVILA---NVCMKSPL 72
T+ TERL L+ +EL + + ++ SD E YM V +A+ ++ ++ ++
Sbjct: 3 TLYTERLTLRKMELEDADVLCQYWSDPEVTKYMNITPFTDVSQARDMIQMINDLSLEGQA 6250 Query: 73 GIYAMIEKESQKMIGIIIELEIRDEFS--AEFGYILNKNYNGKGYMTEACSKLMSIGFEHL 130
+++I KE+ ++IG + D+ + AE GY L +N+ GKG+ +EA KL+ GF L
Sbjct: 63 NRFSIIVKETDEVIGTCGFNMIDQENGRAEIGYDLGRNHWGKFASEAVQKLIDYGFTSL 12255 Query: 131 DLERIYARFDINNNKKSGNVMERIGMKKEGELRHLAKNPGEWKTRAYYSILKEEY 185
+L RI A+ + N S ++ + +KEG LR K KG +S+LK EY
Sbjct: 123 NLNRIBAKVEPENTPSIKLILNSLSFQKEGLLRDYEK-AKGRLIDVYMFSLLKREY 176

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 209

A DNA sequence (GBSx0223) was identified in *S.agalactiae* <SEQ ID 669> which encodes the amino acid sequence <SEQ ID 670>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.5180 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA87001 GB:Z46902 unknown [Saccharomyces cerevisiae]
  Identities = 105/224 (46%), Positives = 148/224 (65%), Gaps = 3/224 (1%)

  Query: 1 MGDVVENFTEGKNPKIDTLNGKTVRIEKINPD-HFEDLFQVYGELSTEDSLTYISFSKFN 59
          +G VE +T P+ L G T R+E ++ + H +LF Y E + TY+ F
20  Sbjct: 11 VGADVEGWTTTRAFPEKVVLKGNTCRLEPLDRERHGSELFSSAYSEAG-QKLWTYLPAGPFT 69

  Query: 60 SKNEFDVFFQTLLKSEDPPYYLAIVDNNNTGKVLGTFSLMRIDTKNRVVMGWVVYSSKLKQ 119
          + E+ F + L +++D AI++ T + +GT L+RID N +E+G+VV+S +L++
  Sbjct: 70 NLEEYLEFIKELNKETAVGTLCLIRIDEANGSLEVGVYVVFSPELQK 129

25  Query: 120 TRIATEAQYLVMKYVFEELCYRRYEWKCDSDLNAPSNNSAKRLGFTFEGTFRQAVVYKGRN 179
          T IATEAQ+L+MKYVF++L YRRYEWKCDSLN PS +A RLGF +EGTFRQ VVYKGR
  Sbjct: 130 TIIATEAQFLLMKYVFDLQYRRYEWKCDSDLNGPSRRAAMRLGFKYEGTFRQVVVYKGR 189

  Query: 180 RDTNWYSILDKEWPEKKTRFEKWLDLDSNFAVNGYQIRSLSSEQ 223
          RDT W+SI+DKEW + FE+WLD +NF NG Q R +++I +
  Sbjct: 190 RDTQWFSIIDKEWLIRKTFEELDKTNFE-NGKQKRGIAIRE 232
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 210

A DNA sequence (GBSx0224) was identified in *S.agalactiae* <SEQ ID 671> which encodes the amino acid sequence <SEQ ID 672>. Analysis of this protein sequence reveals the following:

```
40 Possible site: 39
  >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood =-12.15 Transmembrane 25 - 41 ( 20 - 49)

----- Final Results -----
45  bacterial membrane --- Certainty=0.5861 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

50 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8519> and protein <SEQ ID 8520> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -3.31
 GvH: Signal Score (-7.5): -4.44
 Possible site: 39
 5 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -12.15 threshold: 0.0
 INTEGRAL Likelihood = -12.15 Transmembrane 25 - 41 (20 - 49)
 PERIPHERAL Likelihood = 11.94 59
 modified ALOM score: 2.93
 10 *** Reasoning Step: 3

----- Final Results -----
 bacterial membrane --- Certainty=0.5861 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 672 (GBS43) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 4; MW 34kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 9; MW 58kDa) and in Figure 15 (lane 20 4; MW 59kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 211

25 A DNA sequence (GBSx0225) was identified in *S.agalactiae* <SEQ ID 673> which encodes the amino acid sequence <SEQ ID 674>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> May be a lipoprotein
 ----- Final Results -----
 30 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9519> which encodes amino acid sequence <SEQ ID 9520> 35 was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 212

A DNA sequence (GBSx0226) was identified in *S.agalactiae* <SEQ ID 675> which encodes the amino acid sequence <SEQ ID 676>. Analysis of this protein sequence reveals the following:

Possible site: 44
 >>> Seems to have no N-terminal signal sequence
 45 INTEGRAL Likelihood = -1.54 Transmembrane 165 - 181 (164 - 181)
 INTEGRAL Likelihood = -0.85 Transmembrane 67 - 83 (67 - 84)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1617 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA82211 GB:Z28353 similar to a B.subtilis gene (GB:
      BACHEMEHY_5) [Clostridium pasteurianum]
5   Identities = 40/185 (21%), Positives = 87/185 (46%), Gaps = 6/185 (3%)

Query: 18 MPKGKQKVILSAIELFASQGFHGTSTAQLAKNAEVSQATIYKYFETKDKLLVFILELIVQ 77
        M K K + SAI++F++ G++G + ++A NA V++ T+Y +F++K+++ +I+E V
      Sbjct: 1 MNKTKDNIFYSAIKVFSNNNGYNGATMDEIASNAGVAKGTLYYHFKSKEIFKYIIIEGVN 60
10
Query: 78 TIGRPFFTELSTFSTKEELIHFFVQDRFKFIEKNNDLIKILMQELLINSETSTIFTKLIN 137
        + E + + + I KN D K++ +L ++
      Sbjct: 61 LMKNEIDEATDKEKTALEKLKAVCRVQLNLiyKNRDFFKVIASQLWGKELRQLELRDIMR 120
15
Query: 138 STDPNITKIFNCLSEGNSL---NKMEILRAVIGQFITFFIQLY-ILNIKPENLEELKQI 193
        + +I + E S+ N + + A +G + + LY ++N + +N+ ++ +
      Sbjct: 121 NYVVHIEEFVKDAMEAGSIKKGNSLFVAYAFLGTLCS--VSLYEVINAENDNINNTIENL 178
20
Query: 194 EKQIL 198
        IL
      Sbjct: 179 MNYIL 183
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
25 vaccines or diagnostics.

Example 213

A DNA sequence (GBSx0227) was identified in *S.agalactiae* <SEQ ID 677> which encodes the amino acid sequence <SEQ ID 678>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
30  >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2389 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
35
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
40 vaccines or diagnostics.

Example 214

A DNA sequence (GBSx0228) was identified in *S.agalactiae* <SEQ ID 679> which encodes the amino acid sequence <SEQ ID 680>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
45  >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood ==-13.32 Transmembrane 341 - 357 ( 333 - 361)
      INTEGRAL Likelihood ==-10.93 Transmembrane 253 - 269 ( 238 - 277)
      INTEGRAL Likelihood ==-10.77 Transmembrane 172 - 188 ( 166 - 196)
      INTEGRAL Likelihood = -8.01 Transmembrane 225 - 241 ( 215 - 251)
      INTEGRAL Likelihood = -7.01 Transmembrane 21 - 37 ( 18 - 42)
      INTEGRAL Likelihood = -2.66 Transmembrane 285 - 301 ( 283 - 301)

----- Final Results -----
```

-281-

bacterial membrane --- Certainty=0.6328 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB42664 GB:AL049819 putative integral membrane protein
  [Streptomyces coelicolor A3(2)]
  Identities = 60/156 (38%), Positives = 101/156 (64%), Gaps = 1/156 (0%)
10 Query: 176 LMGFMVFFFVFLISGMALLKERTSGTLDRLLATPVKRSDIVFGYMLSYGILAIQTIVIV 235
          L+G      +FL++ +A L+ERTSGTL+RLA P+ +D++ GY L++G LAI+Q+ +
Sbjct: 77 LLGIFPLITMFLVTISIATLERTSGTLERLLAMPLKGKDLIAGYALAFGALAIQVSALAT 136
15 Query: 236 LSTIWLLDIQVVGSIFSIIIVNFILALVALSLGIIMSTIAKSEFQMMQFIPLIIMPQLFF 295
          +W L+ V GS+ +++V+ AL+ +LG+ +S A SEFQ +OF+P +I PQL
Sbjct: 137 GLAVWFGLDVTGSPWLPLLVALLDALLGTALGLFVSAFAASEFQAVQFMPAVIFPQLLL 196
Query: 296 SGII-PLENMASWAQTVGKILPLSYSGDALTKIIMY 330
          G+ P+NM + V +LP+SY+ D+ ++++
20 Sbjct: 197 CGLFTPRDNMHPALEAVSDVLPMSYAVDGMNEVLRH 232
```

There is also homology to a DNA sequence which was identified in *S.pyogenes* <SEQ ID 681> which encodes the amino acid sequence <SEQ ID 682>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
25 >>> Seems to have no N-terminal signal sequence
    INTEGRAL Likelihood = -11.41 Transmembrane 263 - 279 ( 246 - 284)
    INTEGRAL Likelihood = -7.70 Transmembrane 231 - 247 ( 224 - 258)
    INTEGRAL Likelihood = -4.99 Transmembrane 20 - 36 ( 18 - 39)
    INTEGRAL Likelihood = -3.72 Transmembrane 349 - 365 ( 345 - 368)
30     INTEGRAL Likelihood = -3.45 Transmembrane 187 - 203 ( 182 - 204)

----- Final Results -----
      bacterial membrane --- Certainty=0.5564 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
35      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB12662 GB:Z99108 similar to ABC transporter (ATP-binding
  protein) [Bacillus subtilis]
  Identities = 92/369 (24%), Positives = 180/369 (47%), Gaps = 25/369 (6%)
40 Query: 12 IKRKKTSYVTFFLMPILTTLALSLSFSNNNQAKIGILDKDNSQISKQFIAQLKQNKKYD 71
          I +K +Y+ F P+L T+ S+ N+++ ++ I+D+D++ +S+ +I QLK +
Sbjct: 15 IFKKPQNYLIMFAAPLLLTFVFGMSLSGNDDKVRLAIVDQDDTILSQHYIRQLKAHDDMY 74
45 Query: 72 IFTKIKKEHIDHYLQDKSLEAVLTIDKGFSDKVLQGKSQKLNRSIANSEITEWVKAQTN 131
          +F + + L+ K+ ++ I+ F ++ +GK +L R VK
Sbjct: 75 VFENMSESKASEKLKQKKIAGTIIVISRSFQTQLEKGKHPELIFRHGPELSEAPMVQYAE 134
50 Query: 132 YLLENYNIIGDVALGNEDTFNRL-----ILQKNQQLNYYDVVKQVTLDRSRSKAVSST 182
          L NI A T +K++ + V + TL+D+ S T
Sbjct: 135 SALATLNIQVTAKTASQTAGENWKAAYKTVFAKKHEDIVPAVTRQTLSDKKEGAEASDT 194
Query: 183 TT---GFLLILMLGSTSIVYSGILADKSSQLYHRLMLSRLFR---YMLSYVCVGFVA 235
          + GF ++ ++ + IL + + ++ RL+ +++SR Y+LS+ +G++
55 Sbjct: 195 ASRAAGFSILFVMLTMMGAAGTILEARKNGVWSRLLTASVSRRAEIGAGYVLSFFVIGWIQ 254
Query: 236 FTIQIVIMLSLLKVFNISFFVPTSLLLIAIGFGLLIGAITQNSQQSSQLANL 295
          F I ++LS +F I++ P ++++++ LF L +G GL+I A + +Q NL
60 Sbjct: 255 FGI---LLLSTHWLFGINWGNPAAVIVLVS-LFLLTVVGIGLMIAANVRTPEQQLAFGNL 310
Query: 296 IVMPTSMLAGCLWPLSITPSYMQAIGKLLPQNWLDSAIA-IFQSGGTLSQAWPYLLALMG 354
          V+ T M++G WP+ I P +MQ+I + LPQ W +S + I +G ++ +L + G
Sbjct: 311 FVIATCMVSGMYWPIDIEPKFMQSIAEFLPQKWAMSGLTEIIANGARVTD---ILGICG 366
```

-282-

Query: 355 TALALISFS 363
 LA + +
 Sbjct: 367 ILLAFAAIT 375

5

An alignment of the GAS and GBS proteins is shown below:

Identities = 92/375 (24%), Positives = 164/375 (43%), Gaps = 66/375 (17%)

Query: 11 IKELF---RDKRTLAMMFLAPILIMFLMNVMFSANSNTVKVIGTINVNTKVVSNLDNIK 66
 IK LF R K + FL PIL L+ + S ++N + KIG ++ + +S
 Sbjct: 5 IKTLFVLIKRKKTSYVTFFLMPILTT-LLALSLSFNSNNQAKIGILDKDNSQISK---- 58

10

Query: 67 HIQVRSFKFNSSAKKALKSNKIDALISEDNKSYTIVFYANTDSSKTTLT-RQAFKTAVENTM 125
 +F + LK NK + ++ K + Y S + LT + F V
 Sbjct: 59 -----QFIAQ-----LKQNKKYDIFTKIKKEHIDHYLQDKSLEAVLTIDKGFSDKVLQG 107

15

Query: 126 NSKELISQVKILANKNPKLAQSLQTRSKEYIKEKYNY-----GNKNT-----GF 168
 S++L I + N ++ + ++ ++ Y+ E YN GN++T +
 Sbjct: 108 KSQKL---NIRSIANSEITEWVKAQTNYLLENYNIIGDVALGNEDETFNRILQKNQQLNY 163

20

Query: 169 FAKMIPIL-----MGFMVFFFVFLISGM--ALLKERTSGTLDRLLATPVKRSD 214
 K + + GF++ + S + +L +++S RL+ + + R
 Sbjct: 164 DVKQVTLTDRSRSKAVSSTTGFLILMLGSTSVIYSGILADKSSQLYHRLMLSNSLRS-- 221

25

Query: 215 IVFGYMLSY---GILAIIQTTIVLSTIWLIDIQVVGSIIFSVIIVNFIHALVALSLGILM 271
 F YMLSY G +A IVI+LS + + +I ++I+ F + +L+A+ G+L+
 Sbjct: 222 --FRYMLSYVCVGFVAFTIQIVIMLSLLKVFNISFFVPTSLIIFFLFSLLAIGFGLLI 279

30

Query: 272 STLAKSEFQMMQFIFIPLIIMPQLFFSGII-PLENMASWAQTVGKILPLSYSGDALTKIIMY 330
 + ++ Q Q LI+MP +G + PL S+ Q +GK+LP ++ A+ I
 Sbjct: 280 GAITQNSQQSSQLANLIVMPTSMILAGCLWPLSITPSYMOAIGKLLPQNWVLSAIA-IFQS 338

35

Query: 331 GQGLPNVSSNLLVLL 345
 G L LL L+
 Sbjct: 339 GGTLSQLAWPYLLALM 353

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9081> which encodes the amino acid sequence <SEQ ID 9082>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.52 Transmembrane 21 - 37 (17 - 43)
 INTEGRAL Likelihood = -10.30 Transmembrane 351 - 367 (346 - 371)
 INTEGRAL Likelihood = -5.36 Transmembrane 262 - 278 (260 - 285)
 INTEGRAL Likelihood = -2.60 Transmembrane 288 - 304 (288 - 305)
 INTEGRAL Likelihood = -1.81 Transmembrane 229 - 245 (229 - 246)

----- Final Results -----
 bacterial membrane --- Certainty=0.6010(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45

An alignment of the GAS and GBS sequences follows:

Score = 62.5 bits (149), Expect = 9e-12
 Identities = 72/382 (18%), Positives = 166/382 (42%), Gaps = 32/382 (8%)

55

Query: 1 MVLFHLIKESLQIFRNRTALLMMVIFPILMIVILSFAKSSFNTATTVPKLTIRYQLEG 60
 M + + +K ++FR++ L MM + PIL++ +++ F ++ NT + + + ++
 Sbjct: 1 MRITAITEKVIKELFRDKRTLAMMFLAPILIMFLMNVMFSANSNTVKVIGTINVNTKVV 60

60

Query: 61 EKTDYQKNFLAFLKVLNQKLHLETKPSNSLEKDRQRVSEGALTAVLEVKKNQTICKVITNN 120
 L+ H++ + ++ + + A++ + N++ V N
 Sbjct: 61 N-----LDNIKHIQVRSFKFNSSAKKALKSNKIDALIS-EDNKSYTIVFYAN 105

Query: 121 INQQNADLINMLVKNYVDNAKYDSIAALY-----PQQLNHIRKRSVDYVKVSSIQTSK 174

-283-

+ L K V+ + + I+ + P+ ++ RS Y+K + +
 Sbjct: 106 TDSSKTTLTRQAKTAVNTMNSKELISQVKILANKNPKLAQSLQTRS-KYIKE---KYNY 161

5 Query: 175 GMTSADYYA---ISMFTMITFYSMSMSAMNLVLSDRQQRITNRIHLTGVSPLVFGKLI 230
 G + ++A I M M+ F+ + + +L +R +R+ T V S +VFG ++
 Sbjct: 162 GNKNTGFFAKMIPILMGFMVFFFVFLISGMALLKERTSGTLDRLLATPVKRS DIVFGYML 221

10 Query: 231 GAMILATTVQLSLLYIFTRFVLRVNWGTNEWMLIGITASLVYLSVAIGIGLGISIKNEAFL 290
 + +Q ++ + T ++L + + + +I + L +++++GI + K+E +
 Sbjct: 222 SYGILAIIQTIVLSTIWLIDIQVVGSIIFSVIIVNFI ALVALSLGILMSTLAKSEFQ 281

15 Query: 291 TVASNTIIPPIFAFLGGSYVPLTLHSSIINQLSNISPPIKVWNDSLFLYLIFGGQYNP-IPV 349
 II F G +PL + +S + I P+ + D+L +I GQ P +
 Sbjct: 282 MQFIPLIIMPQLFFSG-IIPLENM-ASWAQTVGKILPLS YSGDALTKIIMYGQGLPNVSS 339

Query: 350 TLIVNISIGTIFIIILALIGMRK 371
 L+V + I I + G+++
 Sbjct: 340 NLLVLLLFLIILTIANIFGLKR 361

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 215

A DNA sequence (GBSx0229) was identified in *S.agalactiae* <SEQ ID 683> which encodes the amino acid sequence <SEQ ID 684>. This protein is predicted to be CG1718 gene product (b0794). Analysis of this 25 protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.17 Transmembrane 118 - 134 (117 - 134)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

35 A related GBS nucleic acid sequence <SEQ ID 8521> which encodes amino acid sequence <SEQ ID 8522> was also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 8
 McG: Discrim Score: -10.96
 GvH: Signal Score (-7.5): -4.84
 40 Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -1.17 threshold: 0.0
 INTEGRAL Likelihood = -1.17 Transmembrane 142 - 158 (141 - 158)
 PERIPHERAL Likelihood = 4.98 197
 45 modified ALOM score: 0.73

*** Reasoning Step: 3

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

55 >GP:AAF50837 GB:AE003568 CG1718 gene product [Drosophila melanogaster]
 Identities = 80/204 (39%), Positives = 123/204 (60%), Gaps = 3/204 (1%)

Query: 7 EIIGLIGPSGAGKSTLKI KTMGMEKADKG TALV--LDTQMPDRN ILNQIGYMA QSDALYE 64
 E GL+G +GAGK+T K M G E+ G A V L + +I IGY Q DAL +

-284-

Sbjct: 1394 ECFGLLGNGAGKTTFKMMTGDERISSGAAYVQGLSLESNMNSIYKMICGCPQFDALLD 1453

Query: 65 SLTGLENLFFGKMKGIQKTELKQQITHISKVVVDLENQLDKFVSGYSGGMKRRLSLAIAL 124
LTG E L F ++G+Q++ ++Q ++K +DK YSGG KR+LS AIA+

5 Sbjct: 1454 DLTGREVLRIFCMLRGVQESRIRQLSEDLAKSFGFMKHIDKQTHAYSGGNKRKLSTAIAV 1513

Query: 125 LGNPTVLILDEPTVGIDPSLRRKIWQELINIKDEGHSIFITTHVMDEAE-LTSKVALLLR 183
+G+P+V+ LDEPT G+DP+ RR++W + I+D G SI +T+H M+E E L +++A+++

10 Sbjct: 1514 IGSPSVIYLDEPTTGMDPAARRQLWNMVCRIRDGKSIVLTSHMEECEALCTRLAIMVN 1573

Query: 184 GNIIAFDTPLHLKKQFNVSTIEEV 207
G + HLK +F+ I ++

Sbjct: 1574 GEFKCIGSTQHLKNKFSKGLILKI 1597

Identities = 73/216 (33%), Positives = 128/216 (58%), Gaps = 9/216 (4%)

15 Query: 1 MEVFKGEIIGLIGPSGAGKSTLIKTMLGMKADKGTAALV--LDTQMPDRNILNQIGYMAQ 58
M +F+ EI L+G +GAGK+T I + GM GTA++ D + +G Q

Sbjct: 536 MNMFEDEITVLLGHNGAGKTTISMLTGMFPPTSGTAIINGSDIRTNIEGARMSLGI CPQ 595

20 Query: 59 SDALYESLTGLENLFFGKMKGIQKTELKQQITHISKVVVDLENQLDKFVSGYSGGMKRRL 118
+ L++ ++ ++ FF +MKG++ ++Q++ K+++LE++ + S SGGMKR+L

Sbjct: 596 HNVLFDEMSVSNHIRFFSRMKGLRGKAVEQEVAKYLMIELEDKANVASSKLSGGMKRKL 655

25 Query: 119 SLATALLGNPTVLILDEPTVGIDPSLRRKIWQELINIKDEGHSIFITTHVMDEAE-LTSK 177
S+ AL G+ V++ DEP+ G+DPS RR++W +L+ + G ++ +TTH MDEA+ L +

Sbjct: 656 SVCCALCGDTKVVLCDEPSSGMPSARRQLW-DLLQQEKVGRLLLTHFMDEADVLGDR 714

Query: 178 VALLLRGNIIAFDTPLHLKKQFN----VSTIEEVF 208
+A++ G + T LKKQ+ VS ++ +F

30 Sbjct: 715 IAICMDGELKCQGTSFFLKKQYGSGYRLVSGVQNLF 750

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 685> which encodes the amino acid sequence <SEQ ID 686>. Analysis of this protein sequence reveals the following:

Possible site: 59

35 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.43 Transmembrane 49 - 65 (49 - 65)

----- Final Results -----

bacterial membrane --- Certainty=0.1171(Affirmative) < succ>

40 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAB12660 GB:Z99108 similar to ABC transporter (ATP-binding protein) [Bacillus subtilis]

45 Identities = 151/316 (47%), Positives = 202/316 (63%), Gaps = 18/316 (5%)

Query: 4 VQLTNNVVKSYKNGKKA-VNDVSLISIEAGNIYGLLGPNGAGKSTLINLILGLIPLSSGKIT 62
+Q N+ K+Y GKK V +S S++ G +GLLGPNGAGKST I++I GL+P SG IT

50 Sbjct: 2 LQAENIKKAY--GKKTIVKGISFSLKKGESFGLGPNGAGKSTTISMISGLVPHDSGNIT 59

Query: 63 VLQGS-QKTIRKISSQIGYVPOQDIAYVPLTAYENVELFGSLYGLKGAQLKKQVLKSLEF 121
V G K K +IG VPQ+IA+YP LTA+EN+ +G +YGL + KK+ + LE+

55 Sbjct: 60 VGGYVIGKETAKAKQKIGIVPQEIALYPTLTAEHENLMFWGKMYGLTHDEAKRAAEVLEY 119

Query: 122 VGLHSQAKQFPSQFSGGMKRRNIACALVHSPKLIIFDEPTVGIDPQSRNHILESIRLLN 181
VGL +AK FSMMKRR+NI AL+H P+L+I DEPTVGIDPQSRNHILE+++ LN

Sbjct: 120 VGLTERAKDKIETFSGGMKRRNIAGAALMHKPPELLIMDEPTVGIDPQSRNHILETVKQLN 179

60 Query: 182 KEGATVIYTTHYMEVEALCDYIFIMDHGQVIEEGPKFELEKRYVANLANQITVTL/TDSR 241
+ G TVIYT+HYMEEVE LCD I I+D G++I G K +L R + Q+ V+ +

Sbjct: 180 ETGMTVIYTSHYMEVEFLCDRIGIIDQGEMIAIGTKTDLCRSRLGGDTIIQLTVSGINEA 239

65 Query: 242 HL---ELADKPDWSLIEDGEKMLKIDNSD----MTSVVHQLTQANITFSEIRHNHL 291
L LA D ++ E L LKID S +TS++ + T +I ++

5 Sbjct: 240 FLVAIRSLAHVNDTVHE---LELKIDISAAHHEKVVTSLAEATAHHINLLSLQVQEP 295

Query: 292 NLEEIFLHLTGKKLRD 307
NLE +FL+LTG+ LRD

5 Sbjct: 296 NLERLFLNLTGRTLRD 311

An alignment of the GAS and GBS proteins is shown below:

Identities = 81/211 (38%), Positives = 125/211 (58%), Gaps = 2/211 (0%)

10 Query: 1 MEVFKGEIIGLIGPSGAGKSTLIKTMIGMEKADKGTAALVL-DTQMPDRNILNQIGYMAQS 59
+ + G I GL+GP+GAGKSTLI +LG+ G VL +Q R I +QIGY+ Q
Sbjct: 25 LSIEAGNIYGLLGPNGAGKSTLINLILGLIPLSSGKITVLGQSQKTIRKISSQIGYVPQD 84

15 Query: 60 DALYESLTGLENLFFGKMKGIQKTELKQQITHISKVVVDLENQLDKFVSGYSGGMKRRLS 119
A+Y LT EN+ FG + G++ +LK+Q+ + V L +Q +F S +SGGMKRRRL+
Sbjct: 85 IAVYVPLDTAYENVELFGSLYGLKGAQLKKQVLKSLEFVGLHSQAKQFPSQFSGGMKRRLN 144

20 Query: 120 IAIALLGNPTVLLDEPTVGIDPSLRRKIWQELINKDEGHSIFITTHVMDEAE-LTSKV 178
+A AL+ +P ++I DEPTVGIDP R I + + EG ++ TTH M+E E L +
Sbjct: 145 IACALVHSPKLIIFDEPTVGIDPQSRNHILESIRLLNKEGATVIYTTHMEEVALCDYI 204

25 Query: 179 ALLLRGNIIIAFDTPLHLKKQFNVSTIEEVFL 209
++ G +I L+K++ + ++ +
Sbjct: 205 FIMDHGQVIEEGPKFELEKRYVANLANQIIV 235

SEQ ID 8522 (GBS391) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 74 (lane 7; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 83 (lane 4; MW 55kDa).

GBS391-GST was purified as shown in Figure 217, lane 3.

30 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 216

A DNA sequence (GBSx0230) was identified in *S.agalactiae* <SEQ ID 687> which encodes the amino acid sequence <SEQ ID 688>. Analysis of this protein sequence reveals the following:

35 Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.6732(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

45 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 217

A repeated DNA sequence (GBSx0231) was identified in *S.agalactiae* <SEQ ID 689> which encodes the amino acid sequence <SEQ ID 690>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

-286-

Possible site: 58
 >>> Seems to have an uncleavable N-term signal seq

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP: CAC18596 GB: AJ278419 IS1381 transposase [Streptococcus pneumoniae]
 Identities = 111/129 (86%), Positives = 117/129 (90%)

 15 Query: 1 MKAQAIIVTSQGRIVSLDIAVNCHDMKLFKMSRRNIGQAQAKILADSGYQGIMKMYSQAQ 60
 MK QAIIVTSQGRIVSLDI VNCHDMKLFKMSRRNIGQA KILADSGYQG+MK+Y QAQT
 Sbjct: 1 MKTQAIIVTSQGRIVSLDITVNYCHDMKLFKMSRRNIGQAGKILADSGYQGLMKIYPQAQT 60

 20 Query: 61 PRKSSKLKPPLTLEDKTYNHTLSKERIKVENIFAKVKTFKIFSTTYRNRRKRFGLRMNLIA 120
 RKSSKLKPPLT+EDK NH LSKER KVENIFAKVKTFK+FSTTYR+ RKRFGFLRMNL A
 Sbjct: 61 SRKSSKLKPPLTVEDKACNHASKERSKVENIFAKVKTFKMFSTTYRSHRKRFGLRMNLSA 120

 25 Query: 121 GMINRELGF 129
 G+IN ELGF
 Sbjct: 121 GIINHELGF 129

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 218

A repeated DNA sequence (GBSx0232) was identified in *S.agalactiae* <SEQ ID 691> which encodes the 30 amino acid sequence <SEQ ID 692>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3996 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP: CAC18595 GB: AJ278419 IS1381 transposase [Streptococcus pneumoniae]
 Identities = 110/125 (88%), Positives = 119/125 (95%)

 45 Query: 1 MNYEASKQLTDVRFKRLVGQRTTFEEMLAVLKTAYQRKHAKGGRTPKLSLEDLLMATLQ 60
 MNYEASKQLTD RFKRLVGQRTTFEEMLAVLKTAYQ KHAKGGR PKLSLEDLLMATLQ
 Sbjct: 1 MNYEASKQLTDARFKRLVGQRTTFEEMLAVLKTAYQLKHAKGGRKPKLSLEDLLMATLQ 60

 50 Query: 61 YMREYRTYEQIADFGIHESNLIRRSQWVESTLIQSGFTISKTHLSAEDTVIVDATEVKI 120
 Y+REYRTYE+IAADFG+HESNL+RRSQWVE TL+QSG TIS+T LS+EDTV++DATEVKI
 Sbjct: 61 YVREYRTYEEIAADFGVHESNLLRRSQWVEVTLVQSGVTISRPLSSEDTVMIDATEVKI 120

 55 Query: 121 NRPKK 125
 NRPKK
 Sbjct: 121 NRPKK 125

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 219

A DNA sequence (GBSx0233) was identified in *S.agalactiae* <SEQ ID 693> which encodes the amino acid sequence <SEQ ID 694>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
    INTEGRAL Likelihood = -10.40 Transmembrane 130 - 146 ( 123 - 156)
    INTEGRAL Likelihood = -7.86 Transmembrane 169 - 185 ( 167 - 191)
    INTEGRAL Likelihood = -6.90 Transmembrane 100 - 116 ( 95 - 118)
    INTEGRAL Likelihood = -5.52 Transmembrane 199 - 215 ( 189 - 216)

----- Final Results -----
bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04126 GB:AP001508 unknown conserved protein in others
[Bacillus halodurans]
Identities = 47/207 (22%), Positives = 95/207 (45%), Gaps = 14/207 (6%)

Query: 7 LQKENTLLEGRIDNSNNQTYTDMIVYLRAA-SISPYHQELIRNDIVNMLLEAQERQASLV 65
        L K+N + N + Y D+++Y+R A S S E + +++++ LLEAQ + S
Sbjct: 6 LIKDNNEKRKLLTEENLKVYEDLLLYIRLAHSKSEQETEELLTEELDHILLEAQAKGKSAK 65

Query: 66 SVFGEDRHDFINQVIKSTPKISKKEE-TLQRWDLAILLLTIQMIIIFLGGYLITEALQQSV 124
        +VFG++ + +++I PK+ KE L + L++ T+ ++F G Y + V
Sbjct: 66 AVFGDNPQYADEIIGEIPKMTKERFGLFAYGLSMFFATV--LVFSGIYRMLRYYVFQV 123

Query: 125 PDLPITLLDVLFIAFISIIAVKIADTIIYATYNFDK---SKEKKYFFRYIFLILSLII 180
        + + + A+ +I ++ IA ++ + + + K F +I + +I
Sbjct: 124 GEAVSEVYVGT--ALITTIASIVIAWMFVFFFQYFRWSCFRINKVFEFFILWLGGMIP 181

Query: 181 AYILIGKYYHLP---FINIPLWIYLI 203
        + Y P I IP+++Y +
Sbjct: 182 FALFFFALLYFTPNVGRMIEIPVYLYFV 208
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 220

A DNA sequence (GBSx0234) was identified in *S.agalactiae* <SEQ ID 695> which encodes the amino acid sequence <SEQ ID 696>. This protein is predicted to be minor extracellular protease epr precursor (epr).

Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL Likelihood = -10.72 Transmembrane 10 - 26 ( 5 - 33)

----- Final Results -----
bacterial membrane --- Certainty=0.5288(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

-288-

A related GBS nucleic acid sequence <SEQ ID 8523> which encodes amino acid sequence <SEQ ID 8524> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1 Crend: 8
McG: Discrim Score: 12.11
5 GyH: Signal Score (-7.5): -4.02
    Possible site: 29
    >>> Seems to have an uncleavable N-term signal seq
ALOM program count: 1 value: -10.72 threshold: 0.0
10      INTEGRAL Likelihood = -10.72 Transmembrane 8 - 24 ( 5 - 33)
      PERIPHERAL Likelihood = 13.74 219
modified ALOM score: 2.64

```

*** Reasoning Step: 3

```

15 ----- Final Results -----
    bacterial membrane --- Certainty=0.5288 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20 !GB:Z99123 extracellular serine protease [Bacillus s...
    >GP:CAB15866 GB:Z99123 extracellular serine protease [Bacillus subtilis]
    Identities = 44/150 (29%), Positives = 80/150 (53%), Gaps = 14/150 (9%)
    Query: 37 QMDTVESSVNHSQDSQLTEAQDMLDKFEKKPSEKLLKDVELALNKLNSSSKKEALQKRFK 96
            ++D V+S N + +A+D + K EK +++ + +A+NKL N + K+ LQKR
    Sbjct: 428 RL DKVQSYRN-----VKDAKDKVAKAEKYKTQQTVDTAQTAINKLPNGTDKKNLQKRLD 481
    Query: 97 KAKDKYLKDEADRKATKDATDLVIEILEQAPSEENVLKAAAVNKLTVKESKEALQKRIDT 156
            + K +Y+ A+K A D V E++ + +V A++A+ KL K +LQKR++
    Sbjct: 482 QVK -RYI-----ASKQAKDKVAKAEKSKKKTDVDSAQS AIGKL PASSEKTS LQKRLNK 533
    Query: 157 VKTQYGLIGNQTPSSSVAETTEQGTANPAS 186
            VK+ Q+ S++ ++T+ A S
    Sbjct: 534 VKSTNLKTAQQSVSAEKKSTDANAAKAQS 563
    Identities = 39/124 (31%), Positives = 64/124 (51%), Gaps = 2/124 (1%)
    Query: 35 TTQMDTVESSVNHSQDSQLTEAQDMLDKFEKKPSEKLLKDVELALNKLNSSSKKEALQKR 94
            +++ + + +N V + L AQ + EKK ++ + A+N+L K ALQKR
    Sbjct: 521 SSEKTS LQKRLNKVKSTNLKTAQQSVSAEKKSTDANAAKAQS A V N Q L QAGKD K T ALQKR 580
    Query: 95 FKKAKDKYLKDEADRKATKDATDLVIEILEQAPSEENVLKAAAVNKLTVKESKEALQKRI 154
            K K K EA K T A V+ E+ + + + + A++AVN+L K LQKR+
    Sbjct: 581 LD KV KKKVAAA EAKK VETAKAK--VKKA EKD KTKSKTS AQS A V N Q L KAS NEK T K L QKRL 638
    45 Query: 155 DTVK 158
            + VK
    Sbjct: 639 NAVK 642

```

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 697> which encodes the amino acid sequence <SEQ ID 698>. Analysis of this protein sequence reveals the following:

```

    Possible site: 41
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL Likelihood = -4.99 Transmembrane 24 - 40 ( 23 - 43)
55 ----- Final Results -----
    bacterial membrane --- Certainty=0.2996 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

60 The protein has homology with the following sequences in the databases:

```
>GP:CAB15866 GB:Z99123 extracellular serine protease [Bacillus subtilis]
```

-289-

Identities = 43/130 (33%), Positives = 71/130 (54%), Gaps = 8/130 (6%)

Query: 41 GSHPQTQDKVA---KHSKSAASLLKKAVKAVNDADRLATAAAIQEAKAVDKLAESSKKK 97
 G P + +K + + +K ++ LK A ++V+ A++ +T A +AQ AV++L K

5 Sbjct: 516 GKLPASSEKTSQKRLNPKVNSTNLKTAQQSVSAAEKKSTDANAAKAQSAVNQLQAGKDGT 575

Query: 98 TLQEQLN-----VAKAKQEQEQAATQAVKAAEETLNQNLDIAQKAVNDLSNKGKKAALQ 152
 LQ++L+ VA A+ ++ + A VK AE+ + K AQ AVN L +K LQ

Sbjct: 576 ALQKRLDKVKKVAAAEEAKKVTAKAKVKAEEKDKTKSKTSQAQSAVNQLKASNEKTKLQ 635

10 Query: 153 SRLDAILPAK 162
 RL+A+ P K

Sbjct: 636 KRLNAVPKKK 645

Identities = 31/105 (29%), Positives = 53/105 (49%), Gaps = 1/105 (0%)

15 Query: 54 SKSAASLLKKAVKAVNDADRLATAAAIQEAKAVDKLAESSKKTLQEQLNVAKAKQEQE 113
 +++ S A +AV A++ I +A++ + +L S K L + +L+ ++ + +
 Sbjct: 380 AQATDSAYAAEQAVKKAEQTKAQIDINKARELISQLPNSDAKTALHKRLDKVQSYRNVK 439

20 Query: 114 DAATQAVKAAEETLNQNLDIAQKAVNDLSNKGKKAALQSRLDI 158
 DA + KA E+ Q D AQ A+N L N K LQ RLD +
 Sbjct: 440 DAKDKVAKA-EKYKTQQTVDTAQTAINKLPNGTDKKNLQKRLDQV 483

An alignment of the GAS and GBS proteins is shown below:

25 Identities = 61/233 (26%), Positives = 115/233 (49%), Gaps = 13/233 (5%)

Query: 2 SMKIDKKELLALIASIILLIFASVTFFLFKDHGTTQMDTVESSVNHVSQSQLTEAQDMLD 61
 SM +KE L + S++ + + +F H TQ + S + + S L +A ++

30 Sbjct: 12 SMTKSQKEALYWMLSVLTTILIGGSCLIFGSHPQTQDKVAKHSKS--AASLLKKAVKAVN 69

Query: 62 KFEKKPSEKLLKVELALNLNSNSKKEALQKRFKKAKDKYLKDEADKKATKDATDLVEI 121
 ++ + +++ + A++KL+ SSKK+ LQ++ AK K ++++A AT V+
 Sbjct: 70 DADRLATAAAIQEAKAVDKLAESSKKKTQEQQLNVAKAKQEQE-----ATQAVKA 122

35 Query: 122 LEQAPSEENVLKAEAAVNKLTVKESKEALQKRIDTVKTQYGLIGNQTPSSVAETTEQGT 181
 E+ ++ A+ AVN L+ K K ALQ R+D + +I ++ P S E T+
 Sbjct: 123 AEETLNQNLDIAQKAVNDLSNKGKKAALQSRLDAILPAKPII-DEFPRQS-GEITDNSY 180

40 Query: 182 ANPASQDTSSYVNQNVAPTYE-QPQANNTPVTPGVNNTP-TPGTGTVPATNG 232
 P D S + + +PT + + + + VTP ++ P P T + P+ +G
 Sbjct: 181 WTPFPGDVSDTYDNSQSPTLDPSSSESSASDVTQPSPHPDPIPQPTSSEPSDSG 233

SEQ ID 8524 (GBS278) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 6; MW 40kDa).

45 The GBS278-His fusion product was purified (Figure 206, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 305), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 221

A DNA sequence (GBSx0235) was identified in *S.agalactiae* <SEQ ID 699> which encodes the amino acid sequence <SEQ ID 700>. Analysis of this protein sequence reveals the following:

Possible site: 53

>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1466 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-290-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 5 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 222

A DNA sequence (GBSx0236) was identified in *S.agalactiae* <SEQ ID 701> which encodes the amino acid sequence <SEQ ID 702>. This protein is predicted to be N-acetylglucosamine-6-phosphate deacetylase 10 (nagA). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
```

```
15 bacterial cytoplasm --- Certainty=0.4607 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9297> which encodes amino acid sequence <SEQ ID 9298> 20 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG21688 GB:AY007718 N-acetylglucosamine-6-phosphate deacetylase
      [Lactococcus lactis subsp. cremoris]
      Identities = 113/178 (63%), Positives = 135/178 (75%)
25 Query: 131 GIYFEGPYFTEEYKGAQNPIYMRNPNEEFAQWQKAAGLITKIALAPEREGVEEFVSAI 190
      GI+FE GP+F TEE KGAQNP YMR+ + E WQ+AA G++ KI LAPEREG E+F+
      Sbjct: 1 GIFFEGPFFTEEKKGAQNPKYMRDAKMWELEDWQEAHGMKKI LAPEREG SED FIRKA 60
30 Query: 191 TKQGVIVALGHNSNGTYKEAKKAVKAGASVWHAYNGMRLTHREPGMVGVAVYNLPNTYAE 250
      T+ GV +ALGHSM TYK+A V+AGASVWH +NGM G+TH+EPGMVGA+ N PNTYAE
      Sbjct: 61 TESGVVIALGHNSNATYKQAVAGVQAGASVWHTFNGMSGMTHQEPMVGAILNTPNTYAE 120
35 Query: 251 LICDGHVDPVACDILMTQKGHNHVALITDCMAAGGAPGDYMLGELPVVSNGTARL 308
      LICDGHV P A +I++ KG +HV LITD M A G PDG YMLGE V V +G A L
      Sbjct: 121 LICDGHVVRPEAAEIVVKMKGADHVVLTDSMRAAGLPDGPYMLGEYEVEVRDGAWL 178
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 703> which encodes the amino acid sequence <SEQ ID 704>. Analysis of this protein sequence reveals the following:

40 Possible site: 40
>>> Seems to have no N-terminal signal sequence

```
----- Final Results -----
```

```
45 bacterial cytoplasm --- Certainty=0.3114 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 227/300 (75%), Positives = 262/300 (86%)
50 Query: 9 MTKYIKADRFFYADHVKGNGYLEIKDNHFGKWIENISGQEEILDYSGYQIAPGLVDTHIH 68
      MT Y+KAD F+Y V+ GYL + D FG+W E + +I+DY+GYQIAPGLVDTHIH
      Sbjct: 1 MTCYLIKADCFFYYPTEVPRPAGYLSLHDGVFGEWTEIVPADAQIIDYTGYQIAPGLVDTHIH 60
```

-291-

```

Query: 69 GFAGADVMDCDSEGILRMSAGLLSTGVTSFLPTTLTSDTKLEEASKSVAAVAGKEQGAK 128
       G+AGADVMD ++GI +MS GLL+TGVTSFLPTTLTS ++LE+ S ++A+VA + +GAK
      Sbjct: 61 GYAGADVMDNAQGIHQMQSEGILLATGVTSFLPTTLTSTFEQLEKVGSTIASVADQVKGAK 120

5   Query: 129 IGGIYFEGPYFTEEYKGAAQNPIYMRNPNLEEFQAQWQKAAGKLITKIALAPEREGVEEFVS 188
       IGGIYFEGPYFTEEYKGAAQNP YM+ P LEEF WQKAAGGLI KIALAPER+GV+EFVS
      Sbjct: 121 IGGIYFEGPYFTEEYKGAAQNPSYMKTPLREFDAWQKAAGKLICKKIALAPERDGKVKEFVS 180

10  Query: 189 AITKQGVTVALGHNSNGTYKEAKKAVKAGASVWHAYNGMRLTHREPGMVGAVYNLPNTY 248
       A+TKQGVTVALGHNSNGTY+EAK+AV+AGASVWHAYNGMRLTHREPGMVGAVYNLPNTY
      Sbjct: 181 AVTKQGVTVALGHNSNGTYQEAKEAVQAGASVWHAYNGMRLTHREPGMVGAVYNLPNTY 240

15  Query: 249 AELICDGHHVDPVACDILMTQKGHNHVALITDCMAAGGAPDGDYMLGELPVVSNGTARL 308
       AELICDGHHV P+ACDILM QKGH+HVA+ITDCM AGG+PDGDY+LGE VVV+NGTARL
      Sbjct: 241 AELICDGHHVSPPIACDILMQQKGHDHVAMITDCMRAGGSPDGDYLLGEFSVVVANGTARL 300

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 223

20 A DNA sequence (GBSx0237) was identified in *S.agalactiae* <SEQ ID 705> which encodes the amino acid sequence <SEQ ID 706>. Analysis of this protein sequence reveals the following:

```

Possible site: 25
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3709 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

30 A related GBS nucleic acid sequence <SEQ ID 9307> which encodes amino acid sequence <SEQ ID 9308> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:CAB16112 GB:Z99124 yyaQ [Bacillus subtilis]
  Identities = 40/110 (36%), Positives = 62/110 (56%), Gaps = 12/110 (10%)
35
Query: 121 IAKTFEDSVDYPFAKHPQYASYRVSG--KWWALLFPLKMKGKLENVPAQLSED---EVEVL 175
       + + + S DYP+ K+P YAS R + KWW L+ + +P +L D E+++L
      Sbjct: 11 VKEKYGTSPDYPWEKYPNYASLRHTSNKKWYGLIMNV-----LPEKLGLDGHGEIDIL 63

40  Query: 176 NIKVNPQDMIEILQKEGIYPSYHMSKKTWSIVLDNTLS DIEIFKLVSDS 225
       N+K P+ + L E I P YHM K+ W+SIVL+ T + EI+ L+ S
      Sbjct: 64 NLKCPPEISDRRLNRGENILPGYHMDKEHWISIVLERTDPEGEIYNLIEQS 113

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 707> which encodes the amino acid sequence <SEQ ID 708>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2541 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

55  Identities = 114/247 (46%), Positives = 169/247 (68%), Gaps = 1/247 (0%)

Query: 7 MSIESDFRKKRFIFSSLEEFGFIFKSDQEQYIYCQTFMDNDFKAIITISLDGKIAGKVIDS 66

```

-292-

MS+ +D+F ++ I L +GF K D Y Y + FM+ +F+A + I G I +VID
 Sbjct: 1 MSLATDYFSRQTPIVEKLMAYGFEKRDNGYFYNERFMEGEFEAQLRIDEAGNIWDRVIDC 60

Query: 67 ALEEEYLPLRAANYNGSFVGEVRSAYMAILGDISDSCCKDLLFTKDQSRLAEKIAKTFE 126
 LEE+YLPL+ A +G++ G+VR+AY+ +L +S +C + F Q+NRLA+ I K +
 Sbjct: 61 DLEEDYLPLQQAAWQGTYTGQVRAAYLELLERLSVACFEATPFQSMQANRLAKHITKEWS 120

Query: 127 DSVDYPFAKHPQYASYRVSGKWYALLFPLKMGKLENVPAOLSEDEVEVLNIKVNPQDMIEI 186
 D +DYPF KHP A+YRV GKWA++F L KL+ +P +L EV+ +KVNP+
 Sbjct: 121 DPMDYPFEKHPDLATYRVGGKWYAMIFSLLADKLQIPERLVGQTCEVMTVKVNPKAFAQ 180

Query: 187 LLQKEGIYPSYHMSKKTWVSIVLDNTLSDIEIFKLVSDSRKLVSHNKKSN-SEPEFWIIP 245
 LLQ+EGIYP+YHMSKK W+SI+LD+ ++D +++ LV+ SR+LV+ N SN + P++W+IP
 Sbjct: 181 LLQOEGIYPAYHMSKKNWISIILDDKVTDDKLWTLVTQSRQLVNPNGLSNPNGPDYWVIP 240

Query: 246 ANPKFYD 252
 AN K+YD
 Sbjct: 241 ANLKYYD 247

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 224

A DNA sequence (GBSx0238) was identified in *S.agalactiae* <SEQ ID 709> which encodes the amino acid sequence <SEQ ID 710>. This protein is predicted to be transposase for insertion sequence element is905.

Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1824 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9601> which encodes amino acid sequence <SEQ ID 9602>
 was also identified.

A related GBS nucleic acid sequence <SEQ ID 9595> which encodes amino acid sequence <SEQ ID 9596>
 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA25167 GB:L20851 transposase [Lactococcus lactis]
 Identities = 325/391 (83%), Positives = 365/391 (93%)

Query: 12 MTQFTTELLNFLAQKQDIDEFRSSLETAMNDLLQVELSAFLGYEPYDKAGYNTGNSRNG 71
 MTQFTTELLNFLAQKQDIDEFFR+SLETAMNDLLQ ELSAFLGYEPYDK GYN+GNSRNG
 Sbjct: 1 MTQFTTELLNFLAQKQDIDEFRTSLETAMNDLLQAEELSAFLGYEPYDKVGYNNSGNSRNG 60

Query: 72 AYTRRFETKYGVVNLLIPRDRNGEFSPALIPSYGRDNHLEEMVIKLYRTGVTTREISDI 131
 +Y+R+FETKYG V L IPRDRN+P+YGRD+HLEEMVIKLY+TGVTTREISDI
 Sbjct: 61 SYSRQFETKYGTQQLSIPRDRNGNFSPALLPAYGRRDDHLEEMVIKLYQTGVTTREISDI 120

Query: 132 IERMIGHHYSPATVSNISKATQENVASFHERSLEANYTVLYLDGTYLPLRRGTVSKECIH 191
 IERMIGHHYSPAT+SNISKATQENVAF+FHRSLEANY+VL+LDGTYLPLRRGTVSKECIH
 Sbjct: 121 IERMIGHHYSPATNISKATQENVATFHERSLEANYSVLFQDGTLPRLRGTVSKECIH 180

Query: 192 IALGVTSYGHKAILGYDIAPNENNAsWSDLLERFKGQGVOOVSLLVSDGFNGLQLIQQA 251
 IALG+T G KA+LGY+IAPNENNAsWS LL++ + QG+QQVSLVV+DGF GL+Q+I QA
 Sbjct: 181 IALGITPEGQKAVLGYEIAPNENNAsWSTLLDKLQNQGIIQQVSLVVDGFKGLEQIISQA 240

Query: 252 FPMAKQQRCLVHIGRNIASKVKRADRALILEQFKTIYRAINVEEAKQALDSFINEWKPHY 311

-293-

+P+AKQQRCL+HI RN+ASKVKRADRA+ILEQFKTIYRA N+E A QAL++FI EWKP Y
 Sbjct: 241 YPLAKQQRCLIHISRNLLASKVKRADRAVILEQFKTIYRAENLEMAVQALENFIAEWKPKY 300

5 Query: 312 KKVIETLESIENLLIFYEFPHQIWGSIYSTNLIESLNKEIKRQTKKVVFPNEESLERYL 371
 +KV+E+LE+ +NLL FY+FP+QIW SIYSTNLIESLNKEIKRQTKKV+FPNEE+LERYL
 Sbjct: 301 RKVMESELNTDNLLTFYQFPYQIWHSIYSTNLIESLNKEIKRQTKKVLFPPNEEALERYL 360

10 Query: 372 VTLFSDYNFKQGQRIHKGFGQCTDTLESLFD 402
 VTLF DYNFKQ QRIHKGFGQC DTLESLFD
 Sbjct: 361 VTLFEDYNFKQSQRHKGFGQCADTLESLFD 391

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 711> which encodes the amino acid sequence <SEQ ID 712>. Analysis of this protein sequence reveals the following:

Possible site: 15
 15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3054 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 20 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 111/128 (86%), Positives = 122/128 (94%)

25 Query: 12 MTQFTTELLNFLAQKQDIDEFRSSLETAMNDLLQVELSAFLGLYEPEYDKAGYNTGNSRNG 71
 MTQFTTELLNFLAQKQDIDEFRSSLE AMNDLLQVELSAFLGLYEPEY+K GYNTGNSRNG
 Sbjct: 1 MTQFTTELLNFLAQKQDIDEFRSSLEIAMNDLLQVELSAFLGLYEPEYKEGYNTGNSRNG 60

30 Query: 72 AYTRRFETKYGVVNLLIPRDRNGEFSPALIPSYGRRDNHLEEMVIKLYRTGVTTREISDI 131
 Y+R+FETKYG+VNL+IPRDRNGEFSP L+PSY RR++HLEE+VIKLY+TGVTREISDI
 Sbjct: 61 TYSRQFETKYGLVNLLIPRDRNGEFSPVLLPSYARREDHLEEIVIKLYQTGVTTREISDI 120

Query: 132 IERMYGH 139
 I+RMYG H
 35 Sbjct: 121 IKRMYGDH 128

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 225

40 A DNA sequence (GBSx0239) was identified in *S.agalactiae* <SEQ ID 713> which encodes the amino acid sequence <SEQ ID 714>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -12.42 Transmembrane 268 - 284 (260 - 286)
 45 INTEGRAL Likelihood = -6.32 Transmembrane 232 - 248 (231 - 254)

----- Final Results -----
 bacterial membrane --- Certainty=0.5967 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD40365 GB:AF036485 hypothetical protein [Plasmid pNZ4000]
 Identities = 69/283 (24%), Positives = 133/283 (46%), Gaps = 9/283 (3%)

55 Query: 11 INVDDLSLQEERF-LPSELLAYARDENESS-FVRDIEGHHLALVYQLLDTQGHVDDVRHVP 68
 IN ++ + E+++ + +++ Y D +ES+ +V DI L L D +R++
 Sbjct: 19 INAAEERATLEDQYGIDEDIIEYVTDNDESTNYVYDINEDDQLFIFLAPYALDKDALRYIT 78

-294-

Query: 69 RVIPVTLFLKEDGLFVLANHKNINLVKKALNRV---EKVDSPKHLLLSSLVTAFSKQYFDV 125
 + P + L + LF N I V AL +V S +L + + +
 Sbjct: 79 Q--PFGMLLHKGVLF--NQSGIPEVNTAIYSALDNPEVKSVDAFILETLFTVVVSFIGPI 135

5 Query: 126 LDTISEERDKLINDLRKRPNKSNLARLANLQSGTVHLMGMTKQNFEMLTDLQNIEQDKEN 185
 I+++R+ L L ++ S+L L+ LQ L + N L L
 Sbjct: 136 SRAITKKRNLYDKMLNRKTYNSDLVSLSYLQQTLTFLSSAVQTNLSELDRLPKTHFGVGA 195

10 Query: 186 TRNEKMQLQDAIEARQLSNMCSLNSQVFQELS-SYNNVLSNNLNDNVTTLTIISIGISI 244
 +++ +D IE Q+ M + +QV + + N++ +NNLND + LTI S+ +++
 Sbjct: 196 DQDKIDLFEDVQIEGEQVQRMFEIETQVVDRIDHTLNSLANNNLNDTMKFLTIWSLTMAV 255

Query: 245 IAMVTSFYGMNVKLPFDSDVAVWVLIIILITTTIIMLSIVMYI 287
 +++ FYGMNVKLP + W+L + I+ ++ + + I++ +
 Sbjct: 256 PTIISGFYGMNVKPLAGMQYAWMLTLCGISVVLIVAMLIMLK 298

SEQ ID 714 (GBS422) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 172 (lane 7; MW 60kDa).

GBS422-GST was purified as shown in Figure 219, lane 12.

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 226

A DNA sequence (GBSx0240) was identified in *S.agalactiae* <SEQ ID 717> which encodes the amino acid sequence <SEQ ID 718>. Analysis of this protein sequence reveals the following:

25 Possible site: 45
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.0783 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB61731 GB:AL133220 putative oxidoreductase. [Streptomyces
 35 coelicolor A3(2)]
 Identities = 100/306 (32%), Positives = 152/306 (48%), Gaps = 3/306 (0%)

Query: 3 KVRYGVVSTAKVAPRFTIEGVRLAGNGEVVAVSSRTLESAQAFANKYHLPKAYDKLEDMLA 62
 KVR+G+++T +A RF + + EVVAV+SRT SA+ FA ++ +P+AY E +
 40 Sbjct: 8 KVRWGILATGGMAARFTADLVLDLPDAEVVAVASRTEASAKTFAERFGIPRAYGGWETLAR 67

Query: 63 DESIDVIYVATINQDHYKVAKAALLAGKHVLIVEKPFTLTYDQANELFALAESCNLFMEA 122
 DE +DV+YVAT + H A L AG++VL EKPFTL +A EL ALA +FLMEA
 Sbjct: 68 DEDVDVYYVATPHSAHRTAAGLCLEAGRNVLCFKPTLNAREAAELVALARENGVFLMEA 127

45 Query: 123 QKSVFIPMTQVIKKLLASGEIGEVEVISISSTAYPN-IDHVTWFRELELGGGTVHFMAPYA 181
 P+ + +K+L+A G IGEV S+ + R+ GGG + + Y
 Sbjct: 128 MMWCNCPLVRRLLKELVADGAIGEVRSQADFGLAGPFPAAHRLRDPAQGGGALLDLGVYP 187

50 Query: 182 LSYLQYLFDATITHASGTATFPKGQS DSQSKLLLQLSNGVLVDIFLTTRLNLPHEMIIYG 241
 +S+ Q L T + A + D Q+ LL N L I + P+ I G
 Sbjct: 188 VSFAQQLLGEP-TDVAARAVLSEEGVDLQTGALLSYGNALASIHCSCITGGTPNSASITG 246

55 Query: 242 TEGRLIIPH-FWKITHAKLVRNDTSARTIQVDMVSDFEKEAYHVSQMILEGQRVSHIMTP 300
 +EGR+ +P+ F+ H L R + + D + H ++ ++ R +P
 Sbjct: 247 SEGRIDVPGNGFFFDPDHFVLHRTGRDPQEPRADGPRESLRHEAEEVMRALRAGETESP 306

Query: 301 QLTLSG 306
 + L G

Subjct: 307 LVPLDG 312

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

5 Example 227

A DNA sequence (GBSx0241) was identified in *S.agalactiae* <SEQ ID 721> which encodes the amino acid sequence <SEQ ID 722>. This protein is predicted to be valyl-tRNA synthetase (valS). Analysis of this protein sequence reveals the following:

Possible site: 36
 10 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.00 Transmembrane 794 - 810 (794 - 810)

 ----- Final Results -----
 15 bacterial membrane --- Certainty=0.1001(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA57558 GB:L08854 valyl-tRNA synthetase [Lactobacillus casei]
 20 Identities = 543/881 (61%), Positives = 679/881 (76%), Gaps = 12/881 (1%)

 Query: 5 LSPKYNPAEVEEGRYQTWLDQDVFKPSGDTEAKPYSIVIPPPNVTGKLHLGHAWDTTLQD 64
 L+PKY+ VEEGRYQ WLD+DVFVKPSGD +AKPYSIVIPPPNVTGKLH+GHAWDTTLQD
 Subjct: 27 LAPKYDHKAVEREGRYQEWLDEDVFKPKPSGDKKAKPYSIVIPPPNVTGKLHMGHAWDTTLQD 86
 25
 Query: 65 IIIRQKRMQGFDTLWLPGMHDAGIATQAKVEERLREQGISRYDLGREKFLDKVWEWKDEY 124
 I+IRQKRR++GFDTLWLPGMHDAGIATQAKVE +LR++GISRYDLGREKF+ KVWEWKDE+
 Subjct: 87 IVIRQKRIEGFDTLWLPGMHDAGIATQAKVEAKLRKEGISRYDLGREKFVQKVWEWKDEF 146

 Query: 125 AATIKSQWGMGLSVDYSRERFTLDEGLSKAVRKVFVDLYNKGIYRGEFIINWDPAART 184
 A TI QW KMGLS+DYSRERFTLD+GL++AVR+VFVDLYN+G IYRGE+I+NWDP ART
 Subjct: 147 AKTIHGQWAKMGLSLDYSRERFTLDKGGLNQAVRRVFVDLYNQGLIYRGEYIVNWDPQART 206

 Query: 185 ALSDIEVIHKDVEGAFYHMNYMLEDGSRALEVATTRPETMFQDVAVAVNPEDARYKDLIG 244
 ALSDIEVIHKD +GAFYH+ Y DGS +E+ATTRPETM GD AVAV+P D RYKD++G
 Subjct: 207 ALSDIEVIHKDDKGAFYHVVKYPFADGSGYIEIAATTRPETMMGDTAVAVHPGDERYKDMVG 266

 Query: 245 QNVILPIINKPIPVADEHADPEFGTGVVKITPAHDPNDFAVGQRHNLQPVNVMDGGTM 304
 +ILP+ N+ IPI+ D + DPEFGTG VKITPAHDPNDF VG RH+L ++N MNDDGTM
 Subjct: 267 TELILPLANRKIPIIEDAYVDPEFGTGAVKITPAHDPNDFQVGNRHDLKRINTMNDGTM 326

 Query: 305 NELADEFNGMDRFEARKAVVAKLESIGLNLVKIKKTTHSVGHERTGVVVEPRLSTQWFVK 364
 NE A ++ GMDRFEARKA+VA L+ G L+K++ HSVGHERTGV VE RLSTQWFVK
 Subjct: 327 NENAGKYQGMDRFEARKAMVADLDKAGLLLKVEPIVHSVGHERTGVQEARLSTQWFVK 386

 Query: 365 MDQIYAKNAI-ANQDTEDKVIFYPPRFNDTFMSWMENVHDWVISRQLWWGHQI PAWYN-VN 422
 M LA+ AI A Q+ + KV F P RF T++ WMEN+HDWVISRQLWWGHQI PAWYN
 Subjct: 387 MKPLAAAIKAQQEPDKKVTFVPERFEHTYLQWMENIHDWVISRQLWWGHQI PAWYNQQT 446

 Query: 423 GEMYVGEDAPEG-DGWTQDEDVLDLTFSSALWPFPSTMGMWPDTTEADF KRYFPTSTLVGTG 481
 GE YVG +AP+ + W QD DVLDLTFSSALWPFPSTMGMWP+T+A D+KRY+PT TLVTGT
 Subjct: 447 GETYVGMEA PKDIENWKQDPDVLDLTFSSALWPFPSTMGMWPNTDAPDYKRYYPTDTLVGTG 506

 Query: 482 DIIFFWVSRMIFQSLEFTGRQPFNSVNLIHGLIRDEEGRKMSKSLGNGIDPMVDVIEKYGAD 541
 DII FWV+RMIFQ L FT ++PF LIHGL+RDE+GRKMSKSLGNGIDPMVDVIEKYGAD
 Subjct: 507 DIIPFWVARMIFQGLHFTHQRPFQYTLIHGLMRDEQGRKMSKSLGNGIDPMVDVIEKYGAD 566

 Query: 542 ALRWFLSNGSAPGQDVRFSYEKMDASWNFINKIWNISRYILMNNNEGILTLDQARENVEKVV 601
 ALRWFL G+ PGQD RFSY+++A+WNFINKIWNISR++MN L Q +
 Subjct: 567 ALRWFLITGNKPGQDTRFSYKQVEAAWNFINKIWNISRFVMMLGDLDTPOQQPD----- 620

-296-

Query: 602 NSQVGNVTDRWILHNLNETVGKVTEENFDKFEFGVAGHILYNFIWEEFANWYVELTKEVLY 661
 +++D+W+ LNET+ +V + +FEFG G LYNF W A+WYVE++KEVLY
 Sbjct: 621 -PSTFDLSDKWLFAQLNETIKQVMQLSARFEGMGRFLYNFTWNVLADWYVEMSKEVLY 679

5 Query: 662 SDNEDEKVITRSVLLYTLQDQILRLHPIMPVTEEIF--GQYAEGSIVLASYPQVNATFE 719
 D+E K R L Y LDQILRLHP+MPFV +++ + SIV ASYP N FE
 Sbjct: 680 GDDEQAKAAKRVNLAYALDQILRLHPVMPFVHGKLWLALPHTGKSIVTASYPVANTAFE 739

10 Query: 720 NQTAHKGVESLKDILRSVRNSRAEVNVAPSXPITILVKTSDELESFFKDNNSNYIKRFTN 779
 N A +++ LIR VR R E + ILVK +D L+ F+ N ++I RF N
 Sbjct: 740 NADATSAMDAIIAIIRGVVRGIRKEAGAPLTKVDILVVLTDPAKPIFEQNPDFIDRFVN 799

15 Query: 780 PETLEISSLATAPELAMSSVITGAEIFLPLADLLNVEEELARLEKELEAKWQKELDMVGKK 839
 + + + +A P++A S+VITGA IF+PL +L++++EE A+L K+ K ++E+ + KK
 Sbjct: 800 SKAFTVGTDVAEPKMAGSAVITGATIFVPLNELIDLDEEKAKLTKDAKLEQEIARIDKK 859

20 Query: 840 LSNERFVANAKPEVVQKEKDKQTDYQTKYDATIARIEEMKK 880
 L+N+ F++ A VV +++ K++D++ + +T R+E++++
 Sbjct: 860 LNNQGFLSKAPEAVVAEQRTRKRSDFEDQLTSTKQRLEQLQR 900

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 723> which encodes the amino acid sequence <SEQ ID 724>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.5062 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below:

Identities = 782/878 (89%), Positives = 818/878 (93%)

35 Query: 4 ELSPKYNPAEVEEGRYQTWLDQDFVKPGSGDTEAKPYSIVIPPPNVTGKLHLGHAWDTTLQ 63
 ELSPKYNPAEVE GRYQ WLD DVFKPSGD +AKPYSIVIPPPNVTGKLHLGHAWDTTLQ
 Sbjct: 3 ELSPKYNPAEVEAGRYQKWLDAADVFKPGSDQAKPYSIVIPPPNVTGKLHLGHAWDTTLQ 62

40 Query: 64 DIIIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLREQGISRYDLGREKFLDKVWEWKDE 123
 DIIIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLREQGISRYDLGR+KFLDKVWEWKDE
 Sbjct: 63 DIIIRQKRMQGFDTLWLPGMDHAGIATQAKVEERLREQGISRYDLGRDKFLDKVWEWKDE 122

45 Query: 124 YAATIKSQWGKMGLSVDYSRERFTLDEGLSKAVRKVFVLDLYNKGIYRGEFIINWDPAAR 183
 YA TIK QWGKMGLSVDYSRERFTLDEGLSKAVRKVFVLDLY NKGIYRGEFIINWDPAAR
 Sbjct: 123 YATTIKEQWGKMGLSVDYSRERFTLDEGLSKAVRKVFVLDLYKKGIYRGEFIINWDPAAR 182

50 Query: 184 TALS DIEVIHKDVEGAFYHMNYMLEDGSRALEVATTRPETMFGDVAVAVNPEDARYKDLI 243
 TALS DIEVIHKDVEGAFYHMNYMLEDGSRAL+VATTRPETMFGDVAVAVNPED RYKDLI
 Sbjct: 183 TALS DIEVIHKDVEGAFYHMNYMLEDGSRALQVATTRPETMFGDVAVAVNPEDPRYKDLI 242

55 Query: 244 GQNVILPIINKPPIVADEHADPEFGTVVKITPAHDPNDFAVGQRHNLPOQVNVMNDG 303
 G+NVILPI+NK IPIV DEHADPEFGTVVKITPAHDPNDF VGQRHNLPOQVNVMNDG
 Sbjct: 243 GKNVILPIVNKLPIVGDEHADPEFGTVVKITPAHDPNDFEVGQRHNLPOQVNVMNDG 302

60 Query: 304 MNE LADEFNGMDRFEARKAVVAKLESLGNLVKIKKTHSVGHSER TGVV EPRLSTQWFV 363
 MNE LA +F GMDRFEAR+A VAKLE LG LV I+K HSVGHSER+G VVEPRLSTQWFV
 Sbjct: 303 MNE LAGDFAGMDRFEARQATVAKLEELGALVIEKRVHSVGHSERSGAVVEPRLSTQWFV 362

65 Query: 364 KMDQLAQNAIANQDTEDKVEFYPPRFNDTFMSWMENVHDWVISRQLWWGHQIPA WYNNG 423
 KMD+LAK A+ NQ+T+D+V+FYPPRFNDTF+ WMENVHDWVISRQLWWGHQIPA WYN G
 Sbjct: 363 KMD ELAKQAMDQETDDRVDFYPPRFNDTFLQWMENVHDWVISRQLWWGHQIPA WYNAG 422

Query: 424 EMYVGEDAPEGDGTQDEDVLDTWFSALWPFSTMGPWTDEADF KRYFP TISTLVTGYDI 483
 E+YVGE+APEGD WTQDEDVLDTWFSALWPFSTMGPWTDEADF KRYFP TISTLVTGYDI
 Sbjct: 423 EIYVGEAPEGDDWTQDEDVLDTWFSALWPFSTMGPWTDEADF KRYFP TISTLVTGYDI 482

-297-

```

Query: 484 IFFWVSRMIFQSLEFTGRQPFNSVLHGLIRDEEGRKMSKSLGNGIDPMDVIEKYGADAL 543
        IFFWVSRMIFQSLEFTGRQPF NVLHGLIRDEEGRKMSKSLGNGIDPMDVIEKYGAD+L
Sbjct: 483 IFFWVSRMIFQSLEFTGRQPFQNVLIHGLIRDEEGRKMSKSLGNGIDPMDVIEKYGADSL 542

5      Query: 544 RWFLSNGSAPGQDVRFSEYKMDASWNFINKIWNISRYILMNNEGLTLDQARENVEKVNS 603
        RWFLSNGSAPGQDVRFSEYKMDASWNFINKIWNISRYILMNNEGLTLDQARENVEKVNS+
Sbjct: 543 RWFLSNGSAPGQDVRFSEYKMDASWNFINKIWNISRYILMNNEGLTLEDAESNVAKVAAS 602

10     Query: 604 QVGNVTDRWILHNLNETVGKVTEENFDKFEGVAGHILYNFIWEEFANWYVELTKEVLYSD 663
        + GNVTD+WILHNLNET+ KVTEENFDKFEGVAGHILYNFIWEEFANWYVELTKEVLYSD
Sbjct: 603 EAGNVTIDQWILHNLNETIAKVTENFDKFEGVAGHILYNFIWEEFANWYVELTKEVLYSD 662

15     Query: 664 NEDEKVITRSVLLYTLQILRLLHPIMPFTVTEEIFGQYAEGSIVLASYPQVNATFENQTA 723
        NE EKVITRSVLLYTLQILRLLHPIMPFTVTEEIFGQYAEGSIVLASYPQVNATFENQTA
Sbjct: 663 NEAEKVITRSVLLYTLQILRLLHPIMPFTVTEEIFGQYAEGSIVLASYPQVNATFENQTA 722

20     Query: 724 HKGVESLKDLIRSVNSRAEVNVAPSKPITILVKTSDFELESNFKNDSNYIKRFTNPETL 783
        HKGVESLKDLIR+VRN+RAEVNVAPSKPITILVKTSDFELE SNFKNDSNYIKRFTNPETL
Sbjct: 723 HKGVESLKDLIRAVRNARAEVNVAPSKPITILVKTSDFELEDFFFNSNINYIKCFTNPETL 782

25     Query: 784 EISSAIATPELAMSSVITGAEIIFLPLADLLNVEELARLEKELAKWQKELDMVGKKLNE 843
        EISSAIATPELAMSSVITGAEIIFLPLADLLNVEELARLEKELAKWQKELDMVGKKLNE
Sbjct: 783 EISSAIAAPELAMTSIITGAEIYLPLADLLNVEELARLKDKELAQKELDMVGKKLNE 842

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
30 vaccines or diagnostics.

```

Example 228

A DNA sequence (GBSx0242) was identified in *S.agalactiae* <SEQ ID 725> which encodes the amino acid sequence <SEQ ID 726>. Analysis of this protein sequence reveals the following:

```

35    Possible site: 30
        >>> Seems to have no N-terminal signal sequence

        ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.0669 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
40        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 727> which encodes the amino acid sequence <SEQ ID 728>. Analysis of this protein sequence reveals the following:

```

45    Possible site: 57
        >>> Seems to have a cleavable N-term signal seq.

        ----- Final Results -----
        bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.
```

55 An alignment of the GAS and GBS proteins is shown below:

Identities = 148/191 (77%), Positives = 165/191 (85%)

Query: 14 GEKKKMNIIIIIGAQASGKMTIGQEIAKQTGMTLFHNHDSDIFVLRFMPWSPDSIALTESI 73

-298-

```

G + KMN+IIIGAQASGKMTIGQE+A+QTGMTLFHNHDSIDFVLRFMPWS +S AL E I
Sbjct: 3 GAETKMNLIIGAQASGKMTIGQEVARQTGMTLFHNHDSIDFVLRFMPWSQESTALIERI 62

5 Query: 74 RFKFFETFAKTGQEMIFTIVIDFNDSRDVVFLEKIQIVFQSHNQEVLFVELETSLERLK 133
      RF FFETFAKTGQ+MIFTIVIDFND DV LEKIQ VFQS++QEVLFVEL+T++ ERLK
      Sbjct: 63 RFAFFETFAKTGQDMIFTIVIDFNPDVAMLEKIQAVFQSYDQEVLFVELKTDIEERLK 122

Query: 134 RNRTERNLKHKPDKWSESDICSTMDYAFNPEVAPEALTYYHKINNTCLTATEAY 193
      RNRTERNLKHKP KR+I+WSE DI STM YA+FNPE P+ LT+Y KINNT LTA ETA
10   Sbjct: 123 RNRTERNLKHKPDKRNIEWSEQDIQSTMAYAVFNPEEPKTLTHYQKINNTQLTAAETAQ 182

Query: 194 LIIQKINQIKE 204
      LIIQK+ IKE
      Sbjct: 183 LIIQKMTHIKE 193

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 229

A DNA sequence (GBSx0243) was identified in *S.agalactiae* <SEQ ID 729> which encodes the amino acid sequence <SEQ ID 730>. Analysis of this protein sequence reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25      bacterial cytoplasm --- Certainty=0.3614 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

30 >GP:BAB04556 GB:AP001510 unknown conserved protein [Bacillus halodurans]
      Identities = 60/189 (31%), Positives = 102/189 (53%), Gaps = 3/189 (1%)

      Query: 7 EIVDNQLPVVETNRLLRQRKLEDACEIFEFVKLDEVSYPAGFPAVKSLEEEITYIQEIY 66
              E + LP +ET RL LR+ +DA I+++ ++V+ + +S+++ ++ +
      35   Sbjct: 4 EDIYGDPLTETERLRLRKFYKDDAAIYDYASNEQVTKYVLWETHQSIKDSEAFLA--F 61

      Query: 67 PTNLEKEKLPSGYAITLKGDDKVIGSVDFNH-RHEDDIFEIGYLLHPDYWGQGIVPEAAS 125
              N EK S +AI LK ++++IG+VDF + +D E+GY+L YWGQQI+ EA +
      40   Sbjct: 62 ALNKYDEKDVPSPWAIELKRNERMIGTVDFVWWPKDKTAEGLYVLSEPYWGQQGIMTEAVN 121

      Query: 126 ALVEIGFTLLGLHKIELGCGYDYNKQSQAARKLGFTLEANIRDRRDAQGKRCGDMRGFL 185
              ALVE GF + L +I+ C+ N S V K G E R +G + ++
      Sbjct: 122 ALVEIGFTLLGLHKIELGCGYDYNKQSQAARKLGFTLEANIRDRRDAQGKRCGDMRGFL 181

      45   Query: 186 RSEWEKKRR 194
              R ++E+K +
      Sbjct: 182 REDYEQKHQ 190

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 731> which encodes the amino acid sequence <SEQ ID 732>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55      bacterial cytoplasm --- Certainty=0.1864 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

-299-

Identities = 50/58 (86%), Positives = 56/58 (96%)

Query: 137 LHKIELGNCYDYNKQSQAVARKLGFLEANIRDRDAQGKRCGDMRFGLLRSEWEKKRR 194
 LHKIELGNCYDYNKQSQAVARKLGFLEAN RDR+D QG+RCGDMRFGLLRSEWE++++

5 Sbjct: 1 LHKIELGNCYDYNKQSQAVARKLGFLEANARDRKDVQGRRCGDMRFGLLRSEEEQKQ 58

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 230

10 A DNA sequence (GBSx0244) was identified in *S.agalactiae* <SEQ ID 733> which encodes the amino acid sequence <SEQ ID 734>. This protein is predicted to be ribosomal-protein-alanine N-acetyltransferase. Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4066 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 9599> which encodes amino acid sequence <SEQ ID 9600> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:BAB04418 GB:AP001509 ribosomal-protein-alanine
 N-acetyltransferase [Bacillus halodurans]
 Identities = 63/185 (34%), Positives = 95/185 (51%), Gaps = 11/185 (5%)

Query: 53 KALPKLETDLRLILRQRTVGDVPAMFDYVCLEEVAYPAGLSPIASLEDEYDYFENRYQNL 112
 K P LET RLILR+ T D ++ Y+ +EV GL P +LED E +Y++

30 Sbjct: 6 KRFPILETKRLILRKITTDDARSILSYLSDEVMKYFGLEPFQTLEDALG--EIAWYESI 63

Query: 113 EKAKLPMSGYGYGITVKGSRIIGSCAFN-----HRHEDDVFEICYLLHPDYWGHHGYMTEAVA 167
 + +GIT+KG D +IGSC F+ H + FE+ L YWG G +EA+

Sbjct: 64 LHEQTGIRWGITLKGQDEVIGSCGFHQWVPKHHRAEIGFELSKL---YWGQGIASEAIR 119

35 Query: 168 ALIEVGFTLLNLHKIEIRCYDYNKQSRRVAEKLGFITLEATIRDRKDQNQDNRCVNLIYGLL 227
 A+I+ GF L L +I+ N S+R+ EK GF E +R + +Y LL

Sbjct: 120 AVIQYGFEHLELQRQLIEPPNIPSQRLVEKQGFISEGLRSYEYTCGKFDDLYMYSLL 179

40 Query: 228 RSEWE 232
 + +++
 Sbjct: 180 KRDFD 184

There is also homology to SEQ ID 732:

45 Identities = 39/54 (72%), Positives = 44/54 (81%)

Query: 179 LHKIEIRCYDYNKQSRRVAEKLGFITLEATIRDRKDQNQDNRCVNLIYGLL 232
 LHKIE+ CYDYNKQS+ VA KLGFITLEA RDRKD Q RC ++ +GLLRSEWE
 Sbjct: 1 LHKIELGNCYDYNKQSQAVARKLGFLEANARDRKDVQGRRCGDMRFGLLRSEWE 54

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-300-

Example 231

A DNA sequence (GBSx0245) was identified in *S.agalactiae* <SEQ ID 735> which encodes the amino acid sequence <SEQ ID 736>. Analysis of this protein sequence reveals the following:

Possible site: 51
 5 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2719 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 10 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 232

A DNA sequence (GBSx0246) was identified in *S.agalactiae* <SEQ ID 737> which encodes the amino acid sequence <SEQ ID 738>. Analysis of this protein sequence reveals the following:

Possible site: 53
 20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3250 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9597> which encodes amino acid sequence <SEQ ID 9598> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

30 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 739> which encodes the amino acid sequence <SEQ ID 740>. Analysis of this protein sequence reveals the following:

Possible site: 38
 35 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3293 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

40 An alignment of the GAS and GBS proteins is shown below:

Identities = 24/55 (43%), Positives = 38/55 (68%)

Query: 56 LLEGLTANKQDVILKEAGLVSLEAFAKVSEADVLALKGIGPAAIKQLVIDNGVVFAK 110
 ++ G+ ++ + L G+ S +AF + +E D+LALKGIGPA +K+LV+NG F K
 45 Sbjct: 77 VVAGIRSDLVETLYAEGIHSQAQAFKEWTEKDLLALKGIGPATVKKLVENGASFKK 131

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-301-

Example 233

A DNA sequence (GBSx0247) was identified in *S.agalactiae* <SEQ ID 741> which encodes the amino acid sequence <SEQ ID 742>. Analysis of this protein sequence reveals the following:

```

5    Possible site: 25
     >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2901(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
10   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 743> which encodes the amino acid sequence <SEQ ID 744>. Analysis of this protein sequence reveals the following:

```

15   Possible site: 27
     >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2536(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```

25   Identities = 57/84 (67%), Positives = 73/84 (86%)
     Query: 1 MSYEQEFLKDFEELWLQSQISINQMAMDSAKKVLEEDKDERAADAYIRYESKLDAYRFLQG 60
              MSYE+EFLKDFE+W+++QI +NQ+AM ++++V +ED DERA DA+IRYESKLDAY FL G
     Sbjct: 1 MSYEKEFLKDFEDWVKTQIQVNQLAMATSQEVAQEDGDERAKDAFIRYESKLDAYEFLLG 60

30   Query: 61 KFNNYHNQKSFHDLPDGLFGQRHY 84
          KF+NY N K+FHD+PD LFG RHY
     Sbjct: 61 KFDNYKNGKAFHDIPDELFGARHY 84
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 234

A DNA sequence (GBSx0248) was identified in *S.agalactiae* <SEQ ID 745> which encodes the amino acid sequence <SEQ ID 746>. This protein is predicted to be methyltransferase. Analysis of this protein sequence reveals the following:

```

40   Possible site: 61
     >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
45   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 747> which encodes the amino acid sequence <SEQ ID 748>. Analysis of this protein sequence reveals the following:

```

50   Possible site: 35
     >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.3352(Affirmative) < succ>
```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

5 Identities = 26/60 (43%), Positives = 37/60 (61%)

Query: 23 LKNERCPHPKLINVLERKLEIIILGDQKHILEKDSDLISLSPQETHHLRAIENSFKLQIELD 82
 + E P K+I VLE +L L DQK +L ++SLI++ Q+ HHL A + K LQ+ LD
 Sbjct: 42 ISQETSPRDKVILVLEGQLIFDLEDQKQVLTQESLIAIPAQKVHHLEAKTDCKLLQVLLD 101

10

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 235

A DNA sequence (GBSx0249) was identified in *S.agalactiae* <SEQ ID 749> which encodes the amino acid 15 sequence <SEQ ID 750>. This protein is predicted to be integrase (codV). Analysis of this protein sequence reveals the following:

Possible site: 59
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3842 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 236

A DNA sequence (GBSx0250) was identified in *S.agalactiae* <SEQ ID 751> which encodes the amino acid 30 sequence <SEQ ID 752>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> May be a lipoprotein

35 ----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

40 No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 752 (GBS128) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 23 (lane 5; MW 15kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 4; 2 bands).

45 The GBS128-GST fusion product was purified (Figure 198, lane 2) and used to immunise mice. The resulting antiserum was used for FACS (Figure 288), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 237

A DNA sequence (GBSx0251) was identified in *S.agalactiae* <SEQ ID 753> which encodes the amino acid sequence <SEQ ID 754>. Analysis of this protein sequence reveals the following:

```

Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10      bacterial cytoplasm --- Certainty=0.2940(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 755> which encodes the amino acid sequence <SEQ ID 756>. Analysis of this protein sequence reveals the following:

```

Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.2518(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

25 An alignment of the GAS and GBS proteins is shown below:

```

Identities = 30/90 (33%), Positives = 49/90 (54%), Gaps = 10/90 (11%)

Query: 3   TVAVRVDQLKDDATELFQSLGLDMSTAVKMFLIQSVKTQSIPFEIK-----NIKSSV 54
          T+ +RVDD +K A ++ + LG+ MSTA+ MFL Q + T IPF++ N +
30     Sbjct: 15  TLNLRVDDSVKSAADDILKRLGIPMSTAIDMFLNQIILTGGIPFDVSLPEAPQRVNVDYM 74

Query: 55  SDEEFQNLVETKLKGIRVKASDPESVNAFF 84
          S E+F + + T + K +P+ V F+
Sbjct: 75  SQEKFYDKLITSFED--AKTCNPQDVGVKFY 102
```

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 238

A DNA sequence (GBSx0252) was identified in *S.agalactiae* <SEQ ID 757> which encodes the amino acid sequence <SEQ ID 758>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```

Possible site: 24
>>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -2.81 Transmembrane 370 - 386 ( 368 - 388)

45     ----- Final Results -----
      bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9593> which encodes amino acid sequence <SEQ ID 9594> was also identified. A related GBS nucleic acid sequence <SEQ ID 10773> which encodes amino acid sequence <SEQ ID 10774> was also identified.

5 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 759> which encodes the amino acid sequence <SEQ ID 760>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -4.57 Transmembrane 354 - 370 ( 353 - 371)

10 ----- Final Results -----
    bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

15 LPXTG motif: 344-348

An alignment of the GAS and GBS proteins is shown below:

Identities = 64/277 (23%), Positives = 99/277 (35%), Gaps = 31/277 (11%)

```
20 Query: 126 SIGNALPDLPKGTTVAFETPVDTATPGDKPAKVVVTYPDGSKDVTVDVKVVDPRTDADKN 185
      ++ +LP + TT E PV + V + D+ + T P A
Sbjct: 121 AVKDLPPASTESTTQPVEAPVQETQASASDSMVTGDSTSVTTDSPEETPSSESPVAAPALSE 180

25 Query: 186 DPAGKDQQNVNGETPKAEDSIGNLPDKGTTVAFETPVDTATPGDKPAKVVVTYPDGSK 245
      PA Q E P S P T A ETP + A P P + S+
Sbjct: 181 APA---QPAESEEPSVAASSEETPS--PSTPAAPETPEEPAAPSPSPSEEPSVAAPSE 234

30 Query: 246 DTVDVTVKVVDPRTDADKNDPAGKDQQNVNGETPKAEDSIGNLPDKGTTVAFETPVDT 305
      +T P A + PA ++ T + P P + +TP
Sbjct: 235 ETPSPET---PEEPAAPSQPAESEESSVAATTSPS-----PSTPAESET--QTTPAV 281

35 Query: 306 ATPGDKPAKVVVTYPDGSKDVTVDVKVVDPRTDADK-----NDPAGKDQQVNGK 355
      DKP+ P S + TV+ + +DK N + +
Sbjct: 282 TKDSDKPSSAAEK-PAASSLVSEQTQQPTSQRSSDKKEEQEQSYSPNRSLSRQVRAHES 340

Query: 356 GNKLPATGENATPFFNVVALTIMSSVGLLSVSKKKED 392
      G LP+TGE A P F + +T+MS G L V+K++++
Sbjct: 341 GKYL PSTGEKAQPLF-IATMTLMSLFGSLLVTKRQKE 376
```

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 239

A DNA sequence (GBSx0253) was identified in *S.agalactiae* <SEQ ID 761> which encodes the amino acid sequence <SEQ ID 762>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.5289(Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 240

A DNA sequence (GBSx0254) was identified in *S.agalactiae* <SEQ ID 763> which encodes the amino acid sequence <SEQ ID 764>. This protein is predicted to be surface protein Rib. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.06   Transmembrane  39 - 55 ( 39 - 55)
10
----- Final Results -----
      bacterial membrane --- Certainty=0.1426 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
15
```

A related GBS nucleic acid sequence <SEQ ID 9591> which encodes amino acid sequence <SEQ ID 9592> was also identified.

The protein differs significantly from U58333 in several places:

```
Query: 157 TKPDGQVDIVNVSLSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
      T PDG D V+V++ + + DK D K KAED P +G+
20
Sbjct: 683 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQNVGETPKAEDSIGNLPDLPGTTVA 742

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQNV 258
      +D T D K T D + +VT K++ PR DADKNDPAGKDQQNV
25
Sbjct: 743 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQNV 798

Query: 157 TKPDGQVDIVNVSLSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
      T PDG D V+V++ + + DK D K KAED P +G+
30
Sbjct: 841 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQNVGETPKAEDSIGNLPDLPGTTVA 900

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQNV 258
      +D T D K T D + +VT K++ PR DADKNDPAGKDQQNV
Sbjct: 901 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQNV 956

Query: 157 TKPDGQVDIVNVSLSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
      T PDG D V+V++ + + DK D K KAED P +G+
35
Sbjct: 288 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQNVGETPKAEDSIGNLPDLPGTTVA 347

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQNV 258
      +D T D K T D + +VT K++ PR DADKNDPAGKDQQNV
40
Sbjct: 348 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQNV 403

Query: 157 TKPDGQVDIVNVSLSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
      T PDG D V+V++ + + DK D K KAED P +G+
45
Sbjct: 604 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQNVGETPKAEDSIGNLPDLPGTTVA 663

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQNV 258
      +D T D K T D + +VT K++ PR DADKNDPAGKDQQNV
50
Sbjct: 664 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQNV 719

Query: 157 TKPDGQVDIVNVSLSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
      T PDG D V+V++ + + DK D K KAED P +G+
55
Sbjct: 446 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQNVGETPKAEDSIGNLPDLPGTTVA 505

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQNV 258
      +D T D K T D + +VT K++ PR DADKNDPAGKDQQNV
Sbjct: 506 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQNV 561

Query: 157 TKPDGQVDIVNVSLSLTIYNSSALRDKIDEVKK-----KAED-----PKWDEGSRDK 201
      T PDG D V+V++ + + DK D K KAED P +G+
60
```

5 Sbjct: 920 TYPDGSKDTVDVTVKVVDPRTDADKNDPAGKDQQVNGETPKAEDSIGLPDLPKGTTVA 979

Query: 202 VLISLDDIKTDIDNNPK---TQSDIANKITEVTNLEKILVPRIPDADKNDPAGKDQQVN 258
+D T D K T D + +VT K++ PR DADKNDPAGKDQQVN

Sbjct: 980 FETPVDTA-TPGDKPAKVVVTYPDGSKDTVDVT--VKVVDPRT-DADKNDPAGKDQQVN 1035

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 241

A DNA sequence (GBSx0255) was identified in *S.agalactiae* <SEQ ID 765> which encodes the amino acid sequence <SEQ ID 766>. This protein is predicted to be ara-C-like activator. Analysis of this protein sequence reveals the following:

15 Possible site: 30
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.37 Transmembrane 8 - 24 (8 - 25)

----- Final Results -----
20 bacterial membrane --- Certainty=0.1150 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9589> which encodes amino acid sequence <SEQ ID 9590> was also identified.

25 There is homology to SEQ ID 460.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 242

A DNA sequence (GBSx0256) was identified in *S.agalactiae* <SEQ ID 767> which encodes the amino acid sequence <SEQ ID 768>. Analysis of this protein sequence reveals the following:

30 Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35 bacterial cytoplasm --- Certainty=0.1200 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9587> which encodes amino acid sequence <SEQ ID 9588> was also identified.

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 769> which encodes the amino acid sequence <SEQ ID 770>. Analysis of this protein sequence reveals the following:

45 Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.0679 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 135/176 (76%), Positives = 161/176 (90%)

```

5   Query: 1  MSYMKDRQIQKTKVAIYNFISLLQENDYSKITVQDVIGLANGRSTFYSHYESKEVLL 60
      +S M KDRQI+KTK AIY+AFI+LLQ+ +YSKITV+D+I LANVGRSTFY+HYESKE+LL
      Sbjct: 1  VSDMTKDRQIKKTKTAIYSAFIALLQKKEYSKITVRDMITLANVGRSTFYAHYESKEMLL 60

10  Query: 61  KELCEDLFHHLFQGRDVTFEEYLVHILKHFEQNQDSIATLLLSDDPYFLLRFRSELEHD 120
      KELCE+LFHHLF+Q R+VTPE+YLVHILKHFEQN+DSIATLLLS+DPYFLLRF++ELEHD
      Sbjct: 61  KELCEELFHHLFQKRNVTFEDYLVHILKHFEQNQDSIATLLLSDDPYFLLRFPNELEHD 120

Query: 121  VYPRRLREYITKV DIPEDFLKQFLLSSFIETLKWWHLHQ RQKMTVEDLLKYYLT MVE 176
      VYP LR +YI K IPE FLKQF+LSSFIETLKWWHLHQ RQ+M+ +LLKYYL +++
      Sbjct: 121  VYPNLRCKYIDKTTIPEVFLKQFVLSSFIETLKWWHLHQ RQ RMSANELLKYYLELIK 176

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 243

20 A DNA sequence (GBSx0257) was identified in *S.agalactiae* <SEQ ID 771> which encodes the amino acid sequence <SEQ ID 772>. Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have no N-terminal signal sequence

```

```

25 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3573 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

30 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 244

35 A DNA sequence (GBSx0258) was identified in *S.agalactiae* <SEQ ID 773> which encodes the amino acid sequence <SEQ ID 774>. Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
      INTEGRAL    Likelihood = -10.19    Transmembrane 112 - 128 ( 107 - 131)
      INTEGRAL    Likelihood = -8.07    Transmembrane 77 - 93 ( 71 - 97)
      INTEGRAL    Likelihood = -6.10    Transmembrane 144 - 160 ( 138 - 165)
      INTEGRAL    Likelihood = -3.03    Transmembrane 165 - 181 ( 164 - 182)

```

```

40 ----- Final Results -----
      bacterial membrane --- Certainty=0.5076 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 775> which encodes the amino acid sequence <SEQ ID 776>. Analysis of this protein sequence reveals the following:

```

50 Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL    Likelihood = -9.13    Transmembrane 112 - 128 ( 107 - 130)

```

-308-

INTEGRAL	Likelihood = -5.89	Transmembrane	144 - 160 (138 - 163)
INTEGRAL	Likelihood = -5.47	Transmembrane	7 - 23 (6 - 29)
INTEGRAL	Likelihood = -3.50	Transmembrane	77 - 93 (74 - 94)
INTEGRAL	Likelihood = -2.07	Transmembrane	166 - 182 (165 - 183)

5

----- Final Results -----

bacterial membrane --- Certainty=0.4652(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10

An alignment of the GAS and GBS proteins is shown below:

Identities = 212/287 (73%), Positives = 245/287 (84%)

Query: 1 MTSNKKVAIAFILNISFSVLEFFIFGSLFFSGAILADAVHDFGDAIAIGISAITLEKKSKKD 60
 M +KKV I FILN+SFS++EFIFG+LFFSGAILADAVHDFGDAIAIGISA LE+K+ K
 Sbjct: 1 MPASKVTIIFILNLSFSLIEFIFGTLFFSGAILADAVHDFGDAIAIGISAILERKAVKK 60

20

Query: 61 EDTIFSLGYKRFSSLGALITSLILISGSILVMENIPKLWHPPTPVNYHGMFILAVIAIII 120
 E FSLGYKRFSSLGAL T+LILISGS+LVMIE IPKLWHPPT VNY GMF+LA+ AIII
 Sbjct: 61 ESPNFSLGYKRFSSLGALTTNLILISGSLLVMETIPKLWHPPTIVNYDGMFVLAIFAI 120

25

Query: 121 NGLASFILHSGQSKHEEILSLHFLEDILGWLAIIIVISLILNKPLYIILDPLLSVAISTFI 180
 NG ASFI+HS Q+K+EEILSLHFLEDILGWLAII++SLIL WKP YILDPLLS+AI++FI
 Sbjct: 121 NGFASFIIHSNQTNEEILSLHFLEDILGWLAIIILSLILKWKPWYIILDPLLSIAIASFI 180

30

Query: 181 LSKALPKLSTLKLFLDGVPDSIDYAAALHDELKGLSQVRSSINQLNIWSMDGIDNRRAIHC 240
 LSKALPKL++T +FLDGVPDSIDY LH EL L + S+NQLN+WSMDGID+RA IHC
 Sbjct: 181 LSKALPKLVATANIFLDGVPDSIDYCTLHHELSQLPHIVSVNQLNVWSMDGIDHRATIHC 240

35

Query: 241 CLNQLISEKDCRAIRTCQHYKINDVTVEIDYSLREHQNHCKPLKN 287
 CL + +EK CK++IR ICQ Y IN VTVEID SL EHQA+HC L +
 Sbjct: 241 CLRESTTEKHCKKSIRLICQRYNINSVTVEIDTSIENEHQHHCSSLSS 287

40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 245

A DNA sequence (GBSx0259) was identified in *S.agalactiae* <SEQ ID 777> which encodes the amino acid sequence <SEQ ID 778>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.22 Transmembrane 221 - 237 (221 - 237)

----- Final Results -----

bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45

There is also homology to SEQ ID 780.

50

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 246

A DNA sequence (GBSx0260) was identified in *S.agalactiae* <SEQ ID 781> which encodes the amino acid sequence <SEQ ID 782>. Analysis of this protein sequence reveals the following:

Possible site: 13

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.50 Transmembrane 2 - 18 (1 - 18)

55

----- Final Results -----

5 bacterial membrane --- Certainty=0.1999 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
10 vaccines or diagnostics.

Example 247

A DNA sequence (GBSx0261) was identified in *S.agalactiae* <SEQ ID 783> which encodes the amino acid sequence <SEQ ID 784>. This protein is predicted to be dehydrogenase (Zn-dependent). Analysis of this protein sequence reveals the following:

15 Possible site: 15
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -3.77 Transmembrane 171 - 187 (170 - 187)
 20 ----- Final Results -----
 bacterial membrane --- Certainty=0.2508 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

25 >GP:AAG20655 GB:AE005134 alcohol dehydrogenase; Adh2 [Halobacterium
 sp. NRC-1]
 Identities = 169/348 (48%), Positives = 232/348 (66%), Gaps = 9/348 (2%)
 30 Query: 1 MKVATFIEPGKMWITDTPKPVIEQETDAVIKIVRACVCGSDLWWYRGISKRESGSFAGHE 60
 M+ A + PG++ + + PKP IE DAVI++ VCGSDLW+YRG S RE+GS GHE
 Sbjct: 1 MRAAVYQGPGEIAVEEVPKPDIESPEDAVIRVTHTAVCGSDLWFYRGDSDRAGSRVGHE 60
 Query: 61 AIGIVEEVGTVKVTDVSKGDFVIVPFTHCGQCPSCAKGFDGNCTNHQA---AKNVGYQGQ 117
 +GIVEEVG VT V+ GD VI PF CG+C C+ G +C ++ N G QG+
 35 Sbjct: 61 PMGIVEEVGDDVTSTVAPGDRVIAPFAISCCEFCRQGLYTSCVEDESWGSEANGGGQGE 120
 Query: 118 YLRYTNANWALVKIPGQPSDYDNETLNSLLTSDVMATGYHAAATAEVKEGDTVVVMGDG 177
 Y++ A+ LV++P + +D D + L SLL L+DVM TG+HAA +A V EGDT VV+GDG
 Sbjct: 121 YVKCPFADGTLVRVPDRYAD-DEDVLESLLPLTDVMTGHAAVSAGVGEGDTAVVVGDG 179
 40 Query: 178 AVGLCGVIAAKMLGANRIIAMSRHKDRQEIALTFGATDIVERGDEAVKRVLDTNQAGA 237
 AVGLCGV+AA+ LGA RIIAM H+DR ELA FGATD + RGD+A++R DLT+ GA
 Sbjct: 180 AVGLCGVIAAQRQLGAERTIAMIAGHHEDRLEIAAEFGATDTISARGDDAIERARDLTH-GGA 238
 45 Query: 238 DAVLECVGTEQSVDTATQIARPGAVIGRVGIP---QNPDMNTNNLFWKNIQLRGGIASVT 294
 + V+ECVG + +D+A IARPG +G VG+P ++ ++ +F NI +RGG+A V
 Sbjct: 239 NHVMECVGAASAMDSAIIAIARPGGTVGVGVPYGVVEDGGLDVFTMFSNDITIRGGVAPVR 298
 50 Query: 295 TFDKSVLDAVLTHKINPGLVFTKSFVLDDIQKAYEAMDKRDAIKSLV 342
 + + ++ D VL + +P +FTK+ LD + + Y AMD R+AIK LV
 Sbjct: 299 AYAEELMAD-VLQGTLDPSPIFTKTVLDGVPEGYAAMDDREAIKVLV 345

There is also homology to SEQ ID 786.

A related sequence was also identified in GAS <SEQ ID 9145> which encodes the amino acid sequence
55 <SEQ ID 9146>. Analysis of this protein sequence reveals the following:

Possible site: 23

-310-

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.41 Transmembrane 170 - 186

5 ----- Final Results -----
 bacterial membrane --- Certainty=0.3166 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 An alignment of the GAS and GBS proteins is shown below:

Identities = 121/353 (34%), Positives = 182/353 (51%), Gaps = 16/353 (4%)

Query: 1 MKVATFIEPGKMWITDTPKPVIEQETDAVIKIVRACVCGSDLWWYRG-ISKRESGSFAGH 59
 MK AT++ G + + D PKPVI + TDA++++V+ +CG+DL G + + G+ GH
 15 Sbjct: 15 MKAATYLSTGNLQLIDKPKPVIIKPTDAIVQLVKTTICGTDLHILGGDVPACKEGTILGH 74

Query: 60 EAIGIVEEVGTVTDVSKGDFIVPFTHGCGQCPSCKAGFDGNCTNHQAASN---VGYQG 116
 E IGIV+EVG VT+ GD VI+ C C CK G +C + G Q
 Sbjct: 75 EGIGIVKEVGDAVTNFKIGDKVIISCVTSCHTCYYCKRGLSSHCQDGGWILGHLINGTQA 134

20 Query: 117 QYLRYTNAWALVKIPQPSDYDNETLNSLLTLSDVMATGYH-AAATAEVKEGDTVVVMG 175
 +Y+ +A+ +L P D +L+ LSD++ T Y + VK GD V ++G
 Sbjct: 135 EYVHIPHADGSLYHAPDTIDD-----EALVMLSDILPTSYEIGVLP SHVKPGDNVCIVG 188

25 Query: 176 DGAVGLCGVIAAKMLGANRIIAMSRRHKDRQELALTGATDIVERGDEAVKRL-DLTNQ 234
 G VGL ++ + II + ++R E A TFGAT + E VK ++ D+TN
 Sbjct: 189 AGPVGLAALLTVQFFSPANIIIMVDLSQRLEAAKTFGATHHTICSGSSEEVKAIIDDTNG 248

30 Query: 235 AGADAVLECVGTEQSVDTATQIARPGAVIGRGRVGIQPQNP-DMNTNNLFWKNI GLRGGIASV 293
 G D +ECVG + D +I G I VG+ P D N + L+ KNI L G+ +
 Sbjct: 249 RGVDISMECVGYPATFDICQKIIISVGHHIANVGVHGKPVDFNLDELWIKNITLNTGLVNA 308

Query: 294 TTDFDKSVLLDAVLTHKINPGLVFTKSFVLDI IQKAYEAMDKRDAIKSL-VIVD 345
 T + +LL+ + T KI+ + T F L +++KAYE A +L VI+D
 35 Sbjct: 309 NTTE--MLLNVLKTGKIDATRLITHHFKLSEVEKAYETFKHAGANNALKVIID 359

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 248

40 A DNA sequence (GBSx0262) was identified in *S.agalactiae* <SEQ ID 787> which encodes the amino acid sequence <SEQ ID 788>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2169 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD36075 GB:AE001762 hypothetical protein [Thermotoga maritima]
 Identities = 55/128 (42%), Positives = 72/128 (55%), Gaps = 8/128 (6%)

55 Query: 8 IFPKGEKNPYGEFFFQSYLAALAKSPDG--NVSVGNVTFEAGCRNNWHVLDGYQILLV 65
 IF +G K +FF G ++ L +G N V +V FE G R +WH H G QIL+V
 Sbjct: 5 IFERGSKGS-SDFFTGNVWVKMLVTDENGVFTQVYDVFEPGARTHWHSHPGG-QILIV 62

Query: 66 TEGSGWYQEEGKEAVSLKPGDVIVTDKGVRHWGAKKDSEFAHIAITA---GKSEFYEA 121
 T G G+YQE GK A LK GDV+ V HWHGA D E HI I+ G +E+ +
 60 Sbjct: 63 TRGKGFYQERGKPARILKKGDVVEIPPNVVWHGAAAPDEELVHIGISTQVHLGPAEWLGS 122

-311-

```
Query: 122 VSDEEYSR 129
      V++EY +
Sbjct: 123 VTEEEYRK 130
```

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 249

- A DNA sequence (GBSx0263) was identified in *S.agalactiae* <SEQ ID 789> which encodes the amino acid 10 sequence <SEQ ID 790>. This protein is predicted to be gamma-carboxymuconolactone decarboxylase. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4089 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

- 20 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA20070 GB:AL031155 3-oxoadipate enol-lactone
      hydrolase/4-carboxymuconolactone decarboxylase
      [Streptomyces coelicolor A3(2)]
      Identities = 33/93 (35%), Positives = 59/93 (62%), Gaps = 1/93 (1%)
25 Query: 11 QLEEFAPFARYNDDILFGEVWAKEDHLTDKTRSIITISALISGGNLQLEHHLQFAKQN 70
      Q +EF+ +F +     +GE+W +   L   ++RS +T++AL++GG+L++L   HL+ A +N
      Sbjct: 349 QADEFSGDFQEFLTRYAWGEIWRD RPG-LDRRSRSCVT LTALVAGGHLDELAPHILRAALRN 407

30 Query: 71 GVTKEEIADIITHLAFYVGWPKAWSAFNKAKEI 103
      G+T   EI +++   A Y G P A AF A+++
      Sbjct: 408 GLTPGEIKEVLLQAAVYCGVPAANGAFRVAQQV 440
```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 250

A DNA sequence (GBSx0265) was identified in *S.agalactiae* <SEQ ID 791> which encodes the amino acid sequence <SEQ ID 792>. Analysis of this protein sequence reveals the following:

```
40 Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5529 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
45      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 251

A DNA sequence (GBSx0266) was identified in *S.agalactiae* <SEQ ID 793> which encodes the amino acid sequence <SEQ ID 794>. This protein is predicted to be probable transcriptional regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have an uncleavable N-term signal seq

10 ----- Final Results -----
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

15 A related GBS nucleic acid sequence <SEQ ID 9585> which encodes amino acid sequence <SEQ ID 9586> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG08263 GB:AE004901 probable transcriptional regulator
  [Pseudomonas aeruginosa]
20 Identities = 36/148 (24%), Positives = 68/148 (45%), Gaps = 22/148 (14%)

Query: 5   QIVEKPAMILAG-----VTLENVKSQNQEGLQQAIGICKTQPDRFD 45
        +IVE+PA + G           + E+ + + + + GIC QP+ F
Sbjct: 123 RIVERPAFSVVGMEYFGSAPGDTIGQLWERFIPREHEIAGKHDPEVSYGICAQQPNGEFH 182
25 Query: 46   YSATYQVETSVQAPKGLEIIRIPSATYAVISVKGPMPSLQETWRKIIQGFFQENNLIKPA 105
        Y A ++V+ P+G+ ++P+ YAV + KG P + E+++ I       E L+P
Sbjct: 183 YVAGFEVQEGWPVPEGMVRFQVPAQKYAVFTHKGTAP-QIAESFQAIYSHLLAERGLEPK 241

30 Query: 106 NSPNLEIYSSQH--PQDTDYQMEIWLA 131
        + E Y + P D + Q+++++ I
Sbjct: 242 AGVDFEYYDQRFRGPLDPNSQVDLYIPI 269
```

No corresponding DNA sequence was identified in *S.pyogenes*.

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 252

A DNA sequence (GBSx0267) was identified in *S.agalactiae* <SEQ ID 795> which encodes the amino acid sequence <SEQ ID 796>. Analysis of this protein sequence reveals the following:

```
40 Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
    bacterial cytoplasm --- Certainty=0.0887 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB84919 GB:AE000825 conserved protein [Methanothermobacter
  thermoautotrophicus]
50 Identities = 42/130 (32%), Positives = 71/130 (54%), Gaps = 3/130 (2%)
```

-313-

```

Query: 1 MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60
        M+T EM + I +L VAT D +G PN+ P R D++T + +N +T N+ +N
Sbjct: 1 MMTPEMMDAIEKELVFVATADEEGTPNVVPIGFARPLDERTILIAIDNYMKKTIRNLHENP 60

5   Query: 61 KIEIAFVDRERLLGYRFVGTAEIQTEGTYYEAAKKWAEGRMG--VPKAVGIIHVERIFNL 118
        +I + R Y+F GT EI G Y++ +WA+ M PK+ ++ VE I+ ++
Sbjct: 61 RIAL-IPQNARECPYQFKGTVEIFKSGKYFDMVVEWAQNVMTLEPKSAILMTVEEIVSV 119

10  Query: 119 QSGANAGKEI 128
        + G AG+ ++
Sbjct: 120 KPGPEAGEKV 129

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 797> which encodes the amino acid sequence <SEQ ID 798>. Analysis of this protein sequence reveals the following:

```

15 Possible site: 24
      >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0789(Affirmative) < succ>
20      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

25 Identities = 123/128 (96%), Positives = 127/128 (99%)
      Query: 1 MITQEMKEIINSQLAMVATVDAKGQPNIGPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60
              MITQEMK++IN+QLAMVATVDAKGQPNI GPKRSMRLWDDKTFIYNENTDGQTRINIEDNG
      Sbjct: 1 MITQEMKDLINNQLAMVATVDAKGQPNI GPKRSMRLWDDKTFIYNENTDGQTRINIEDNG 60

30      Query: 61 KIEIAFVDRERLLGYRFVGTAEIQTEGTYYEAAKKWAEGRMGVPKAVGIIHVERIFNLQS 120
              KIEIAFVDRERLLGYRFVGTAEIQTEG YYEAAKKWA+GRMGVPKAvgIIHVERIFNLQS
      Sbjct: 61 KIEIAFVDRERLLGYRFVGTAEIQTEGAYYEAAKKWAQGRMGVPKAvgIIHVERIFNLQS 120

35      Query: 121 GANAGKEI 128
              GANAGKEI
      Sbjct: 121 GANAGKEI 128

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 253

A DNA sequence (GBSx0268) was identified in *S.agalactiae* <SEQ ID 799> which encodes the amino acid sequence <SEQ ID 800>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 22
      >>> Seems to have a cleavable N-term signal seq.
      INTEGRAL Likelihood = -5.47 Transmembrane 1028 -1044 (1027 -1048)

----- Final Results -----
      bacterial membrane --- Certainty=0.3187(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
50      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

55 !GB:AF054892 surface antigen BspA [Bacteroides forsy...

>GP: AAC82625 GB:AF054892 surface antigen BspA [Bacteroides

```

forsythus]

Identities = 143/566 (25%), Positives = 243/566 (42%), Gaps = 52/566 (9%)

Query: 95 VPKAKPEVTQEASNNSNDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVG---F 150
 5 +P + + A + + +P TA + L T + T +G F
 Sbjct: 120 IPNSVTTIGEWAFKGCSGLKSITLPNSLTAIGQSALSGCTGLTSITIPNSVTTIGEWAFF 179

Query: 151 SKSGINKLSQTSHLVLPSHAA--DGTQLTQVASFAFTPDKKTAIAEYTSRLGENGKPSRL 208
 10 SG+ ++ + L +A LT + PD T I E + G +G S
 Sbjct: 180 GCSGLTSITFPNSLTAIGESAFYGCGLTSIT--LPDALTTIGESAFK-GCSGLKSIT 234

Query: 209 DIDQKEIIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFDVNKNIAEVNLPESELETISDYAFA 268
 15 + I E ++ LT +T+P+ +IG+ AF + + P SL TI + AF
 Sbjct: 235 FPNSLTTIGESAFYDCGALTSLTPDALTTIGRSASYGCGLKSITFPNSLTTIGESAFY 294

Query: 269 HM-SLKQVKLPDNLKVGELAFFFQNQIGGKLYLPRHLIKLAERAFKSNSRIQTVEFLGSKL 327
 20 + SL + +P++ IG AF+ + LP L + ERAF + + T + + +
 Sbjct: 295 NCGSLTSITIPNSVTTIGRSASYGCGLKSITLPDGTLTIEERAFYNCGVLTTSITIPNSV 354

Query: 328 KVIGEASFQD-NMLRNVMPLPDGLEKIESEAFTGNPGDEHYNNQVVLRTGQNPQLATE 386
 25 IGE++F + L+++ LPDGL IE AF N L + T N E
 Sbjct: 355 ATIGESAFYGCGLKSITLPDGTLTIEWGAFY-----NCGALTSITIPNSVSTIGE 405

Query: 387 NTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVTGFS--NKGLQKVRRNKNLEIPKQH 443
 25 + + +L T D ++ D +++ +++G G + V K ++ K+
 Sbjct: 406 SAFYGC-G-ALKDVTVAWDTPIDIQRD-VFRELTLSGIRLHVPGKKTVEAK--DVWKEF 461

Query: 444 NGITITEIGDNAFRNVDQSKTLRKYDLEEIKLPSTIRKIGAFAFQSNNLKSFEASEDLE 503
 30 N + + G + N D +KTL + P T + + FA ++ L
 Sbjct: 462 NIVEDDDDFGGQLW-NYDAATKTLTITN---PTPDTPKPMPNFATPNDQLW----- 507

Query: 504 EIKEGAFMNNRIGHTLDLKDKLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHL 562
 35 GAF I + + D + +GD AF + + +I LP+SV IG+SAF L
 Sbjct: 508 ---GAFQKE-IQKITIGDGVTSGDFAFSGCDALKSITLPKSVTIGQSAFSGCWDLRS 562

Query: 563 MFIGNKVKTIGEMAFLSNKLESVNLSEQKQLKTIEVQAFS-DNALSEVVLPPNLQTIREE 621
 40 + + + V TIGE AF + LE + + + K + I + F +L+ + LP L I ++
 Sbjct: 563 LTLPDGVNTIGEKAFY-DCLELTSITIPKSFTAIGQETFHCVSLSLTLPDALTAIGKK 621

Query: 622 AF-KRNHLKEVKGSSTLSQITFNAFD 646
 45 AF N L V + + + I NAFD
 Sbjct: 622 AFYSCLNTSVTFPKSITTIGENAFD 647
 Identities = 109/407 (26%), Positives = 175/407 (42%), Gaps = 48/407 (11%)

Query: 222 FNAYQLTKLTIPNGYKSIGQDAFDVNKNIAEVNLPESELETISDYAFAHMS-LKQVKLPDN 280
 50 F+ LT +T+P+ +IG AF + + +P S+ TI ++AF S LK + LP++
 Sbjct: 87 FSDCALTSVTLPSLTAIGDHAFKGCSGLTSITIPNSVTTIGEWAFKGCSGLKSITLPNS 146

Query: 281 LKVIGELAFFDNQIGGKLYLPRHLIKLAERAFKSNSRIQTVEFLGSKLKVIGEASFQD-NN 339
 55 L IG+ A + +P + + E AF T + L IGE++F
 Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTTIGEWAFFGCGLTSITFPNSLTAIGESAFYGCMA 206

Query: 340 LRNVMLPDGLEKIESEAFTGNPGDEHYNNQVVLRTGQNPQLATEVNPDKSLWRA 399
 60 L ++ LPD L I AF G G L++ T N E+ + +
 Sbjct: 207 LTSITLPDALTTIGESAFKGCSG-----LKSITFPNSLTTIGESAFYDCGALTSL 257

Query: 400 TPDMDYTKWLEEDFTYQKNSVTGFSNKGLQKVRRNKNLEIPKQHNGITITEIGDNAFRNV 459
 PD ++T K++ P ++T IG++AF N
 Sbjct: 258 LPD-----ALTTIGRSASYGCGLKSITFPN-----SLTTIGESAFYNC 296

Query: 460 DFQSKTLRKYDLEEIKLPSTIRKIGAFAFQS-NNLKSFEASEDLEEIKEGAFMNNRIGT- 517
 65 L I +P+++ IG AF + LKS + L I+E AF N + T
 Sbjct: 297 G-----SLTSITIPNSVTTIGRSASYGCGLKSITLPDGTLTIEERAFYNCGVLT 347

Query: 518 LDLKDKLIKIGDAAFH-INHIYAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMA 576
 + + + + IG++AF+ + + +I LP+ + I AF GAL + I N V TIGE A
 Sbjct: 348 ITIPNSVATIGESAFYGCGLKSITLPDGTLTIEWGAFYNCGALTSITIPNSVSTIGESA 407

Query: 577 FLS-NKLESVNLSEQKQLKTIEVQAFSDNALSEVVL--PPNLQTIRES 620
 F L+ V ++ + I+ F + LS + L P +T+ E
 Sbjct: 408 FYGCAGALKDVTVAWDTP1-DIQRDVFRELTLSGIRLHVPAGKKTVE 453
 Identities = 111/465 (23%), Positives = 185/465 (38%), Gaps = 56/465 (12%)
 5

Query: 141 VTRGDTLVGFSKGSGINKLSQTSHLVLPSHAADGTQLTQVASFAF-----TPDKKT 190
 +T D L +S S + P+ LT + AF PD T
 Sbjct: 210 ITLPDALTTIGESAFKGCSGLKSITFPN-----SLTTIGESAFYDCGALTSITLPDAL 263

10 Query: 191 AIAEYTSRLGENGKPSRLDIDQKEIIIDEGEIFNAYQLTKLTIPNGYKSIQGDAFVDNKNI 250
 I ++ G +G S + I E +N LT +TIPN +IG+ AF +
 Sbjct: 264 TIGR-SAFYGCGLKSITFPNSLTTIGESAFYNCGLTSITIPNSVTTIGRSAFYGCGL 322

15 Query: 251 AEVNLPESLETISDYAFAHMS-LKQVKLPDNLKVGELAFFDNQIGGGKLYLPRHLIKLAE 309
 + LP+ L TI + AF + L + +P+++ IGE AF+ + LP L +
 Sbjct: 323 KSITLPDGLTTIEERAFYNCGLTSITIPNSVATIGESAFYGCGLKSITLPDGLTIEW 382

20 Query: 310 RAFKSNRIQTVEFLGSKLKVIGEASFQD-NNLRNVMLP-DGLEKIESEAF-----TGNPG 362
 AF + T + + + IGE++F L++V + D I+ + F +G
 Sbjct: 383 GAFYNCGLTSITIPNSVSTIGESAFYGCALKDVTVAWDTPIDIQRDVFRELTLSGIRL 442

25 Query: 363 DEHYNNQVVLRTGQNPHQLATEN-----TYVNPDKSLWRATPDMDYTKWLEEDFTY 415
 + V + + + Y K+L P D K + +F
 Sbjct: 443 HVPAEKVYEAKEVWKEFNIVEDDDFGGLQWNYDAATKTLTITNPTPDTPKPM-PNFAT 501

30 Query: 416 QKNSVTGFSNKGQLQKVRNNKNEIPKQHNGITITEIGDNAFRNVDQSKTLRKYDLEEIK 475
 + + G K +QK+ G +T +GD AF D L+ I
 Sbjct: 502 PNDQLWGAFQKEIQKIT-----IGDGVTSGDFAFSGCD-----ALKSIT 541

35 Query: 476 LPSTIRKIGAFAFQSN-NLKSFEASEDLEEIKEGAFMN-NRIGTLDLKDKLIKIGDAAFH 533
 LP ++ IG AF +L+S + + I E AF + + ++ + + IG FH
 Sbjct: 542 LPKSVTTIGQSAFSGCWDLRSLTLPDGVTNTIGEKAFYDCLELTSITIPKSVTAIGQETFH 601

40 Query: 534 -INHIYAIIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAF 577
 + ++ LP+++ IGE AF AL + + TIGE AF
 Sbjct: 602 YCVSLTSLTLPDALTAIGKKAFYSCNALTSTFPKSITTTIGENAF 646
 Identities = 98/351 (27%), Positives = 152/351 (42%), Gaps = 53/351 (15%)

45 Query: 315 NRIQTVEFLGSKLKVIGEASFQDNNLRNVMLPDLGKIESEAF-TGNPGDEHYNNQVVLRT 374
 ++IQT+ +G + +G +F D L +V LP+ L I AF G G L +
 Sbjct: 68 SKIQTWT-IGDGVTSGVNNAFSDCALTSVTPNSLTAIGDHAFKGCSG-----LTS 117

50 Query: 375 RTGQNPHQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVTGFSNKGQLQVRRN 434
 T P+ + T + S ++ NS+T L
 Sbjct: 118 IT--IPNSVTTIGEWAFKGCSGLKSIT-----LPNSLTAIGQSALSGCTGL 161

55 Query: 435 KNLEIPKQHNGITITEIGDNAF-----RNVDQSKTLRKYD-----LEEIKLPSTI 480
 ++ IP + +T IGE AF + + F + + L I LP +
 Sbjct: 162 TSITIPN-----SVTTIGEWAFFGCSGLTSITFPNSLTAIGESAFYGCALTSITLPDAL 216

60 Query: 481 RKIGAFAFQSN-NLKSFEASEDLEEIKEGAFMN-NRIGTLDLKDKLIKIGDAAFH-INHI 537
 IG AF+ + LKS L I E AF + + ++ L D L IG +AF+ + +
 Sbjct: 217 TTIGESAFKGCSGLKSITFPNSLTTIGESAFYDCGALTSITLPDALTTIGRSAFYGCGL 276

65 Query: 538 YAIVLPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFLS-NKLESVNLSEQKQLKTI 596
 +I P S+ IG SAF G+L + I N V TIG AF + L+S+ L + L TI
 Sbjct: 277 KSITFPNSLTTIGESAFYNCGLTSITIPNSVTTIGRSAFYGCGLKSITLPD--GLTTI 334

70 Query: 597 EVQAFSD-NALSEVVLPPNLQTIREEAFKR-NHLKEVKGSSTLSQITFNAF 645
 E +AF + L+ + +P ++ TI E AF + LK + L+ I + AF
 Sbjct: 335 EERAFYNCGLTSITIPNSVATIGESAFYGCGLKSITLPDGLTTIEWGAF 385
 Identities = 78/282 (27%), Positives = 123/282 (42%), Gaps = 46/282 (16%)

75 Query: 111 NDASKVEVPKQDTASKKETLETSTWEAKDFVTRGDTLVGFSKGSGINKLSQTSHLVLPS-- 168
 N+AS E+P SK +T VT GD + + + + TS + LP+
 Sbjct: 56 NNAS--EIPWHSILQSKIQT-----VTIGDGVTSGVNNAFSDCALTS-VTLPSL 101

80 Query: 169 -----HAADG-----TQLTQVASFAFT-----PDKKTAAEYTSRLGENG 203

-316-

HA G +T + +AF P+ TAI + ++ G G
 Sbjct: 102 TAIGDHAFKGCSGLTSITIPNSVTTIGEWAFKGCSGLKSITLPNSLTAIGQ-SALSGCTG 160

Query: 204 KPSRLIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQDAFDVNKNIAEVNLPELETIS 263
 S + I E F LT +T PN +IG+ AF + + LP++L TI
 Sbjct: 161 LTSITIPNSVTTIGEWAFFGCSCGLTSITFPNSLTAIGESAFYGCALTSITLPDALTTIG 220

Query: 264 DYAFAHMS-LKQVKLPDNLKVGELAFFFDNQIGGGKLYLPRHLIKLAERAFKS-NRIQTVE 321
 + AF S LK + P++L IGE AF+D + LP L + AF + +++++
 Sbjct: 221 ESAFKGCSGLKSITFPNSLTTIGESAFYDCGALTSITLPDALTTIGRSAFYGCGLKSIT 280

Query: 322 FLGSKLVIGEASFQD-NNLRNVMLPDGLEKIESEAFTGNPG 362
 F S L IGE++F + +L ++ +P+ + I AF G G
 Sbjct: 281 FPNS-LTTIGESAFYNCGLTSITIPNSVTTIGRSAFYGCG 321
 Identities = 43/144 (29%), Positives = 70/144 (47%), Gaps = 4/144 (2%)

Query: 220 EIFNAYQ--LTKLTIPNGYKSIGQDAFDVNKNIAEVNLPELETISDYAFAHM-SLKQVK 276
 + + + A+Q + K+TI +G S+G AF + + LP+S+ TI AF+ L+ +
 Sbjct: 505 QLWGAQKEIQKITIGDGVTSGDFAFSGCDALKSITLPKSVTIGQSAFSGCWDLRSLT 564

Query: 277 LPDNLKVGELAFFF DNQIGGGKLYLPRHLIKLAERAFKSNRIFTVEFLGSKLKVIGEASFQ 336
 LPD + IGE AF+D + +P+ + + + F T L L IG+ +F
 Sbjct: 565 LPDGVNTIGEKAFYDCLELTSITIPKSVTIAIGQETFHYC VSLTS LTPDALTAIGKKAFY 624

Query: 337 D-NNLRNVMLPDGLEKIESEAFTG 359
 N L +V P + I AF G
 Sbjct: 625 SCNALTSVTFPKSITTTIGENAFDG 648
 Identities = 43/134 (32%), Positives = 66/134 (49%), Gaps = 12/134 (8%)

Query: 511 MNNRIGTLDLKDKLIKIGDAAFHINHIYAIIVLPESVQEIGRSAFRQNGALHLMFIGNKVK 570
 + + + I T+ + D + +G+ AF + + + LP S+ IG AF+ L + I N V
 Sbjct: 66 LQSKIQTVTIGDGVTSGNNAFSDCALTSVTLPNSLTAIGDHAFKGCSGLTSITIPNSVT 125

Query: 571 TIGEMAFLS-NKLESVNLSEQKQLKTIEVQAFSD-NALSEVVLPPNLQTIREEAFKRNL 628
 TIGE AF + L+S+ L L I A S L+ + +P ++ TI E AF
 Sbjct: 126 TIGEWAFKGCSGLKSITL--PNSLTAIGQSALSGCTGLTSITIPNSVTTIGEWAF----- 178

Query: 629 KEVKGSSTLSQITF 642
 G S L+ ITF
 Sbjct: 179 ---FGCSGLTSITF 189

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 801> which encodes the amino acid sequence <SEQ ID 802>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -2.44 Transmembrane 984 -1000 (984 -1001)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.1977 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

LPXTG motif: 975-979
 An alignment of the GAS and GBS proteins is shown below:
 Identities = 751/1050 (71%), Positives = 861/1050 (81%), Gaps = 45/1050 (4%)
 Query: 3 KKHLKTLALATTVSVVTVSQEVYGLEREESVKQEQTQSA-SEDDWFEEEDNERKTNVSKE 61
 KKHLKT+AL LTTVSVVT++QEV+ L +E +KQ Q S+ S D+ E + K +++
 Sbjct: 2 KKHLKTVALTLLTVSVVTHNQEVFSLVKEPILKQTQASSSISGADYAESSGKSKLKINET 61
 Query: 62 NSTVDETVDSDLFSDGNSNNSSKTESVVSVDPKQVPKAKPEVTQEASNSSNDASKVEVPKQ 121
 + VD+TV+DLFSD + K +Q KA E T E+ S++E K+
 Sbjct: 62 SGPVDDTVTDLFSDKRTTPEKIKDNLAKGPREQELKAVTENT-ESEKQITSGSOLQSKE 120

Query: 122 DTASKKETLETSTWEAKDFVTRGDTLVGFSKSGINKLSQTSHLVLPSHAADGTQLTQVAS 181
 + K TS WE DF+T+G+TLVG SKSG+ KLSQT HLVLPS AADGTQL QVAS
 Sbjct: 121 SLSLNKTVTPSTSNEICDFITKGNTLVGLSKSGVEKLSQTDHVLVLPSQAADGTQLIQQVAS 180

5 Query: 182 FAFTPDKKTAIAEYTSRLGENGKPSRLDIDQKEIIDEGEIFNAYQLTKLTIPNGYKSIGQ 241
 FAFTPDKKTAIAEYTSR GENG+ S+LD+D KEII+E+E+FN+Y L K+TIP GYK IGG
 Sbjct: 181 FAFTPDKKTAIAEYTSRAGENGEISQDVGKEIIINEGEVFNSYLLKKVTIPTGYKHIGQ 240

10 Query: 242 DAFVDNKNIAEVNLPESETISDYAFAHMSLKQVKLPDNLKVGELAFFDNQIGGKLYLP 301
 DAFVDNKNIAEVNLPESETISDYAFAH++LKO+ LPDNLK IGELEAFFDNQI GKL LP
 Sbjct: 241 DAFVDNKNIAEVNLPESETISDYAFAHLLKQIDLPDNLKAIGELEAFFDNQITGKLSLP 300

15 Query: 302 RHLIKLAAERAFKSNRIQTVEFLGSKLKVICEASFQDNNLRNVMLPDGLEKIESEAFTGNP 361
 R L++LAERAFKSN I+T+EF G+ LKVICEASFQDN+L +MLPDGLEKIESEAFTGNP
 Sbjct: 301 RQLMRLAERAFKSNHIKTIEFRGNSLKVVICEASFQDNDLSQLMLPDGLEKIESEAFTGNP 360

20 Query: 362 GDEHYNNQVVLRTTRGQNPHQLATENTYVNPDKSLWRATPDMDYTKWLEEDFTYQKNSVT 421
 GD+HYNN+VVL T++G+NP LATENTYVNPDKSLW+ +P++DYTKWLEEDFTYQKNSVT
 Sbjct: 361 GDDHYNNRVLWTKGKPNPSGLATENTYVNPDKSLWQESPEIDYTKWLEEDFTYQKNSVT 420

25 Query: 422 GFSNKGLQKVRNRNKLLEIPKQHNGITITEIGDNAFRNVDQFSKTLRKYDLEEIKLPSTIR 481
 GFSNKGLQKV+RKNLLEIPKQHNG+TITEIGDNAFRNVDQ+KTLRKYDLEE+KLPSTIR
 Sbjct: 421 GFSNKGLQKVKRNKLLEIPKQHNGVTITEIGDNAFRNVDQFKTLRKYDLEEVKLPSTIR 480

30 Query: 482 KIGAFAFQSNNLKSFEASEDLEEIKEGAFMNNRIGTLDLKDQLIKIGDAAFHINHIYAIV 541
 KIGAFAFQSNNLKSFEAS+DLEEIKEGAFMNNRI TL+LKDQL+ IGDAAFHINHIYAIV
 Sbjct: 481 KIGAFAFQSNNLKSFEASDDLEEIKEGAFMNNRIETLELKDLVIGDAAFHINHIYAIV 540

35 Query: 542 LPESVQEIGRSAFRQNGALHLMFIGNKVKTIGEMAFLSNKLESVNLSEQKQLKTIEVQAF 601
 LPESVQEIGRSAFRQNGA +L+F+G+KVKT+GEMAFLSN+LE ++LSEQKQL I VQAF
 Sbjct: 541 LPESVQEIGRSAFRQNGANNLIFMGSKVKTGEMAFLSNRLEHLDLSEQKQLTEIPVQAF 600

40 Query: 602 SDNALSEVVLPNNLQTIREEAFKRNLKEVKGSSTLSQITFNAFDQNDGDKRGKKVVVR 661
 SDNAL EV+LP +L+TIREEAFK+NHLK+++ +S LS I FNA D NDGD++F KVVV+
 Sbjct: 601 SDNALKEVLLPASLKTIREEAFKKNLKQLEVASALSHIAFNALDDNDGDEQFDNKVVVK 660

45 Query: 662 THNNSHMLADGERFIIDPDKLSSTMVDLEKVLKIIEGLDYSTLRQTTQFREMTTAGKA 721
 TH+NS+ LADGE FI+DPDKLSST+VDLEK+LK+IEGLDYSTLRQTTQFQR+MTTAGKA
 Sbjct: 661 THHNSYALADGEHFIVDPDKLSSTIVDLEKILKLIIEGLDYSTLRQTTQFQFRMTTAGKA 720

50 Query: 722 LLSKSNLQRGEKQKFLQEAQFFLGRVLDKAIAKAEKALVTKKATKNGHLLERSINKAVL 781
 LLSKSNLQRGEKQKFLQEAQFFLGRVLDKAIAKAEKALVTKKATKNG LLERSINKAVL
 Sbjct: 721 LLSKSNLQRGEKQKFLQEAQFFLGRVLDKAIAKAEKALVTKKATKNGOLLERSINKAVL 780

55 Query: 782 AYNNSAIKKANVKRLEKELDLLTDLVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY 841
 AYNNSAIKKANVKRLEKELLLL TVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY
 Sbjct: 781 AYNNSAIKKANVKRLEKELLLL TVEGKGPLAQATMVQGVYLLKTPPLPEYYIGLNVY 840

60 Query: 842 FDKSGKLIYALDMSDTIGEGQKDAYGNPINLVNDEDNEGHTLAVATTADYEGLYIKDILN 901
 FDKSGKLIYALDMSDTIGEGQKDAYGNPINLVNDEDNEGHT LAVATTADYEGL IK ILN
 Sbjct: 841 FDKSGKLIYALDMSDTIGEGQKDAYGNPINLVNDEDNEGHT HALAVATTADYEGLDIKTILN 900

65 Query: 902 SSIDKIKAIRQIPLAKYHRLGIFQAIRAAAEDRLLPKTPKGYLNEVPNYRKQVEKNL 961
 S L ++ +IRQ+P A YHR GIFQAI+NAAAEEA++LLPK
 Sbjct: 901 SKLSQLTSIROVPTAAHYRAGIFQATQNAAAEAEQLLPK----- 939

Query: 962 KPVDYKTPIFNKALPNEKVDGDRAAKGHNINAETNNSVAVTPIRSEQQLHKSQSDVNLPQ 1021
 +++++ + N++ +S + ++ + LP+
 Sbjct: 940 -----PGTHEKSSSESANSKDRG-----LQSNPKTNRGRHSAILPR 977

Query: 1022 TSSKNNFIYEILGYVSLCLLFVLTAGKKGK 1051
 T SK +F+Y ILGY S+ LL L+TA KK K
 Sbjct: 978 TGSKGSFVYIGLYTSSVALLSLITAICKKK 1007

SEQ ID 800 (GBS97) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 12; MW 113.4kDa).

GBS97-His was purified as shown in Figure 193, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 254

- 5 A DNA sequence (GBSx0269) was identified in *S.agalactiae* <SEQ ID 803> which encodes the amino acid sequence <SEQ ID 804>. This protein is predicted to be ribonucleoside-diphosphate reductase alpha chain (nrdE). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4274 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

15 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB96160 GB:AE000050 ribonucleoside-diphosphate reductase alpha
      chain-MPN324 (new), 513 (Himmelreich et al., 1996)
      [Mycoplasma pneumoniae]
20 Identities = 476/725 (65%), Positives = 586/725 (80%), Gaps = 20/725 (2%)

Query: 2 TQSD--AYLSLNAKTRFRDRTGNYHFTSDKEAVEQYMIEHVEPNTMVFTSLIEKLDYLV 59
      TQ D +Y+SLNA T+ F D AVE Y+ EHV+P T VF S E+LD+LV
Sbjct: 12 TQEDELYSISLNAYTKVYG---DFKMDLHHAVEAYIQEHWVKPKTKVFHSTKERLDFLVK 66

25 Query: 60 NNYYESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAYALKTEDNRYYLEHYED 119
      N+YY+ +++ Y+ E +I AYA +F + NFMGA KFYNAYALKT D ++YLE+YED
Sbjct: 67 NDYYDENIIINMYSFEQFEEITRKAYAYRFRYANFMGAFKFYNAVALKTFDGKWLHENYED 126

30 Query: 120 RVVMNALFLAAGDEKAAYDLVDDMLANRFQPATPTFLNAGKRRGEYISCYLLRIEDNME 179
      RVVMN LFLA G+ A L+ ++ NRFQPATPTFLNAG+K+RGE++SCYLLRIEDNME
Sbjct: 127 RVVMNVLF LANGNYNKA KLLKQIITNRFQPATPTFLNAGKRRGEFVSCYLLRIEDNME 186

35 Query: 180 SISRAISTSLQLSKRGGGVALCLTNLREFGAPIKGIKNQATGIVPVMKLLEDSFYANQL 239
      SI RAI+T+LQLSKR GGVAL LTN+RE GAPIK I+NQ++GI+P+MKLLEDGSFYANQL
Sbjct: 187 SIGRAITTTLQLSKRDGGVALLLTNIRESGAPIKKIENQSSGIIPIMKLLEDGSFYANQL 246

40 Query: 240 GQRQGAGAVYLHAAHHPVLTFLDTKRENADEKIRIKSLSLGLVLIPDITFELAKANKDMAL 299
      GQRQGAGAVYLHAAHHPV+V+ FLDTKRENADEKIRIKSLSLGLVLIPDITF LAK N++MAL
Sbjct: 247 GQRQGAGAVYLHAAHHPDVMQFLDTKRENADEKIRIKSLSLGLVLIPDITFTLAKNNEMAL 306

45 Query: 300 FSPYDIERYVGKPMDSISITEYEYETLLANADIRKTFISARKLFQTIAELHFESGYPYILF 359
      FSPYD+ YGK+SDIS+TE Y LLAN I+KTFI+ARK FQT+AELHFESGYPYILF
Sbjct: 307 FSPYDVYEEYGKPLSDISVTEMYYELLANQRIKKTFINARKFFQTVAELHFESGYPYILF 366

50 Query: 360 EDTVNAKNPHKKEGRIVMSNLCEIAQVNTASQFSEDLTFTKVGHDVCCNLGSINIARAM 419
      +DTVN +N H RIVMSNLCEI Q +T S+F DL F KVG+D+ CNLGS+NIA+AM
Sbjct: 367 DDTVNRNAH--PNRIVMSNLCEIVQPSTPSEFHDLAFKKGNDISCNLGSLNIAKAM 424

55 Query: 420 DQAADFEKLTIANSIRALDRVSRSTS DLDAPSIIKKGNAANHAVGLGAMNLHGFLATNHIYY 479
      + F +L+ +I +LD VSR S+L++APSI+KGN+ NHA+GLGAMNLHGFLATN IYY
Sbjct: 425 ESGPEFSELVKLAIESTLDLVSRSVNLETAPSIIQKGNSENHALGLGAMNLHGFLATNQIYY 484

Query: 480 DSQEAIIDFTDCFFYAMAYYAFKASNHLAKEKGTPEGFSESSYADGSYFYQY--TEQNF-E 536
      +S EAIDFT+ FFY +AY+AFKAS+ LA EKG F+ F + +ADGSYF +Y E +F
Sbjct: 485 NSPEAIDFTNIFFYTVAYHAFKASSELALEKGKFKNFENTKFADGSYFDKYIKVEPDFWT 544

60 Query: 537 PKTQRVKNLAAEYGLTLPSEQEDWRKLVQSIKEIGLANAHLLAVAPTGSISYLSCTPSLQ 596
      PKT+RVK L +Y + +P++E+W++L +I++ GLAN+HLLA+APTGSISYLSCTPSLQ
Sbjct: 545 PKTERVKALFQKYQVEIPTRENWKE LALNIQKNGLANSLLAIAPTGSISYLSCTPSLQ 604
```

-319-

Query: 597 PVVSPVEVRKEGALGRVYVPAYKIDADNYVYKKGAYEVGSEAIINIAAAAQKHIDQAIS 656
 PVVSPVEVRKEGALGRVYVPAYKIDADNYVYKKGAYEVGSEAIINIAAAAQKHIDQAIS
 Sbjct: 605 PVVSPVEVRKEGRLGRIYVPAYQLNKDSYPFYKDGAYELGPEPIINIAAAAQQHVDQAIS 664

5 Query: 657 LTLMFTDQATTRDLNKAYIQAFQKQKCASTIYVVRVRQDILEGSYDDMLDDFTSSDLED 716
 LTLMFTD+ATTRDLNKAYI AFK+ C+SIYYVRVRQ++LE SE + + + C
 Sbjct: 665 LTLMFTDKATTRDLNKAYIYAFKKGCSSIYVVRVRQEVLEDSEDH-----TIQMOC 716

10 Query: 717 QSCMI 721
 ++C+I
 Sbjct: 717 EACVI 721

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 805> which encodes the amino acid
 15 sequence <SEQ ID 806>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1843 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

25 >GP: AAC82625 GB: AF054892 surface antigen BspA [Bacteroides forsythus]
 Identities = 124/451 (27%), Positives = 202/451 (44%), Gaps = 65/451 (14%)

Query: 221 FNSYLLKKVTIPTGYKHIGQDAFVDNKNIAEVNLPELETISDYAFAHLA-LKQIDL PDN 279

30 F+ L VT+P IG AF + + + P S+ TI ++AF + LK I LP++
 Sbjct: 87 FSDCALTSVTLPNSLTAIGDHAKGCGSLTSITIPNSVTI GEWAFKGCSGLKSITLPNS 146

Query: 280 LKAIGELAFFDNQITGKLSPRQLMRLAERA-FKSNHIKTIEFRGNLSLKVGAEASFQD-N 337
 L AIG+ A +++P + + E A F + + + I F NSL IGE++F

35 Sbjct: 147 LTAIGQSALSGCTGLTSITIPNSVTI GEWAFFGCGSLTSITF-PNSLTAIGESAFY GCG 205

Query: 338 DLSQLMLPDGLEKIESEAFTGMPGDDHYNNRVLWTKSGKNPSGLATENTYVNPDKSLWQ 397
 L+ + LPD L I AF G G KS P+ L T +S +
 Sbjct: 206 ALTSITLPDALTTIGESAFAKGCSG-----LKSITFPNSLTITIG-----ESAFY 248

40 Query: 398 ESPEIDYTKWLEEDFTYQKNSVTGFSNKGLQKVKRKNLEIPKQHNGVTITEIGDNAFRN 457
 + + + T +++ G S GL K++ P ++T IG++AF N
 Sbjct: 249 DCGALTSPALTTIGRSAFYGCS--GL-----KSITFPN-----SLTTIGESAFTN 295

45 Query: 458 VDFQNKTTLRKYDLEEVKLPSTIRKIGAFAFQS-NNLKSFEASDDLEEIKEGAFMNNRIET 516
 L + +P+++ IG AF + LKS D L I+E AF N + T
 Sbjct: 296 CG-----SLTSITIPNSVTI GRSAFYGCGSLKSITLPDGTLTIEERAFYNCGVLT 346

50 Query: 517 -LELKDKLVTIGDAFH-INHIYAIVLPESVQEIGRSAFRQNGANNLIFMGSKVKTLGEM 574
 + + + + TIG++AF+ + + I LP+ + I AF GA I + + V T+GE
 Sbjct: 347 SITIPNSVATIGESAFTYGCGLKSITLPDGTLTIEWGAFTYNCGALT SITIPNSVSTIGES 406

55 Query: 575 AFLS-NRLEHLDLSEQKQLTEIPVQAFSDNALKEVLL--PASLKTIREEAFKKNHLKQLE 631
 AF L+ + ++ + + I F + L + L PA KT+ E K+ K+
 Sbjct: 407 AFYGC GALKDVTVAWDTP -DIQRDVFRELTLSGIRLHV PAGKKT VYE--AKDWKE-- 460

60 Query: 632 VASALSHIAFNALDDND-GDEQFDNKVVVK 661
 FN ++D+D G Q++ KT
 Sbjct: 461 -----FNIVEDDDFGGLQWNDAATKT 482

An alignment of the GAS and GBS proteins is shown below:

Identities = 534/726 (73%), Positives = 614/726 (84%), Gaps = 5/726 (0%)

Query: 1 MTQS D A-YLSLN AKTRFRDR TGNYHFTSDKEA VEQYMIEH VEPNTMVF TSLIEKLDYLVS 59

-320-

M+Q++A YLSLNA TRF+ G+YHF SDKEAV +Y+ EHV PN M F SL +KL YL++
 Sbjct: 1 MSQTNASYLSLNALTRFKKPDSYHFDSDKEAVRRYLEEHVSPNQMAFNSLEDKLAYLIN 60

5 Query: 60 NNYYESDLLKQYNLEFICQIFEHAYAKKFAFLNFMGALKFYNAYALKTEDNRYYLEHYED 119
 YYE + Y + I + F +AY + + FLN MGA+KFY +YALKT D + YLE +ED
 Sbjct: 61 EGYYEQAIFFDAYPNLILIKEAFHYAYQQGYRFLNLMGAMKFYQSYALKTLDGKQYLETED 120

Query: 120 RVVMNALFLAAGDEKAAYDLVDDMLANRFQPATPTFLNAGKKRGEYISCYLLIEDNME 179
 R VMNALFLA GD+ +D++D +L RFQPATPTFLNAGKKRGEYISCYLLR+EDNME
 10 Sbjct: 121 RAVMNALFLADGIDQTFVFDVIDAILHRRFQPATPTFLNAGKKRGEYISCYLLVEDNME 180

Query: 180 SISRAISTSLQLSKRGGGVALCLTNLREFGAPIKGKIQNQATGIVPVVMKLEDSFSYANQL 239
 SISRAISTSLQLSKRGGGVALCLTNLRE GAPIKGK+I+NQATGIVPVVMKLEDSFSYANQL
 15 Sbjct: 181 SISRAISTSLQLSKRGGGVALCLTNLREIGAPIKGKQATGIVPVVMKLEDSFSYANQL 240

Query: 240 GQRQGAGAVYLHAAHHPVELTFLDTKRENADEKIRIKSLSLGLVIPDITFELAKANKDMAL 299
 GQRQGAGAVYLHAAHHPVELTFLDTKRENADEKIRIKSL+LGLVIPDITF+LAK NKDMAL
 Sbjct: 241 GQRQGAGAVYLHAAHHPVELTFLDTKRENADEKIRIKSLSLGLVIPDITFQLAKENKDMAL 300

20 Query: 300 FSPYDIERYVYGPMSDISITEEYETLLANADIRKTFISARKLFQFTIAELHFESGYPYILF 359
 FSPYDI+R YGK MSDISITEEY+ LLAN I+KT+ISARK FQ IAELHFESGYPY+LF
 Sbjct: 301 FSPYDIKRAYGKDMSDISITEEYDKLLANPAIKKTYISARKFFQLIAELHFESGYPYLLF 360

25 Query: 360 EDTVNAKNPHKKEGRIVMSNLCEIAQVNTASQFSEDLTFTKVGHVDVCNLGSINIARAM 419
 +DTVN +NPH K+GRIVMSNLCEIAQV+T S F EDL+F +G D+CCNLGSINIA+AM
 Sbjct: 361 DDTVNKRNPNAKKGRIVMSNLCEIAQVSTPSTFKEDLSFETIGEDICCNLGSINIAQAM 420

Query: 420 DQAADFEKLIANSIRALDRVSRTSDLAPSIIKGNAANHVGGLGAMNLHGFLATNHIYY 479
 A FE+LI SIRALDRVSR SDL+ APS++ GNAANHVGGLGAMNLHGFLATNHIYY
 30 Sbjct: 421 ADAPHFQLITTSIRALDRVSRVSDLNCAVPSVETGNAANHVGGLGAMNLHGFLATNHIYY 480

Query: 480 DSQEAIIDFTDCFFYAMAYYAFKASNHLAKEKGTFEGFSESSYADGSYFYQYTEQNFEPKT 539
 D++EA+DFTD FF+AMAYYAFKAS LAKEKG F GFS S+Y+DG+YF +Y +++ +P+T
 Sbjct: 481 DTKEAVDFTDLFFHAMAYYAFKASCQLAKEKGAFAGFSLSTYSDGTYFAKYLQEDAKPQT 540

35 Query: 540 QRVKNLLAEYGLTLPSEQDWRIKVQSIKEIGLGLANAHLLAVAPTGSISYLLSSCTPSLQPVV 599
 +V LL +YG TLP+ DW+ LV IK+ GLANAHLLAVAPTGSISYLLSSCTPSLQPVV
 Sbjct: 541 AKVATLLQDYGFTLPTIVADWQALVADIQKQFGLANAHLLAVAPTGSISYLLSSCTPSLQPVV 600

40 Query: 600 SPVEVRKEGALGRVYVPAYKIDADNYVYYKKGAYEVGSEAIINIAAAAQKHIDQAISSLTL 659
 +PVEVRKEG+LGR+YVPAY+ID NY YY++GAYEVG +AII++ AAAQKH+DQAISSLTL
 Sbjct: 601 APVEVRKEGSLGRIYVVPAYQIDQANYAYYERGAYEVGPKAIIDVVAQQKHVDQAISSLTL 660

45 Query: 660 FMTDQATTRDLNKAYIQAFKQKCASIYYYVRVRQDILEGSSESYDD---MLDDFTSSDLED 715
 FMTDQATTRDLN+++YIQAFKQ CASIYYYVRVRQD+L GSE YD+ + +
 Sbjct: 661 FMTDQATTRDLNRSYIQAFKQNCASIYYYVRVRQDVLAGSEQYDEDLSVTAPGASDETTTE 720

Query: 716 CQSCMI 721
 CQSCMI
 50 Sbjct: 721 CQSCMI 726

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 255

55 A DNA sequence (GBSx0270) was identified in *S.agalactiae* <SEQ ID 807> which encodes the amino acid sequence <SEQ ID 808>. This protein is predicted to be nrdI protein (nrdI). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence
 60 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2952 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-321-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

```

5 >GP:AAC71451 GB:U39702 nrdI protein (nrdI) [Mycoplasma genitalium]
    Identities = 77/127 (60%), Positives = 104/127 (81%), Gaps = 1/127 (0%)

10 Query: 7 VVYFSSKSNNNTHRFVQKLACSNQRIPSD-GSSILVPTEDYIILIVPTYAGGGDDTKGAVPKQ 65
      +VYFSS SNNNTHRF++KL ++RIP D SI V+ +Y+LI PTY+GGG+ +GAVPKQ
      Sbjct: 22 IVYFSSISNNNTHRFIEKLGFOHKRIPVDITQSITVSNEYVLICPTYSGGGNQVEGAVPKQ 81

15 Query: 66 VVQFLNVRQNREHCQGVISSGNNTNGDTYAIAGPIIARKLNVPPLLHQFELLGTQEDVTRV 125
      V+QFLN + NRE C+GVI+SGNTNFGDT+ +AG +I++KLNVPPLL+QFELLGT+ DV +
      Sbjct: 82 VIQFLNNKHNRRELCRGVIASGNNTNGDTFCLAGTVISKLNVPLLYQFELLGTKNDVEQT 141

Query: 126 KELLCQF 132
      +++++ F
      Sbjct: 142 QKIIANF 148

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 809> which encodes the amino acid sequence <SEQ ID 810>. Analysis of this protein sequence reveals the following:

```

20 Possible site: 54
    >>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0089 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

30 Identities = 84/125 (67%), Positives = 100/125 (79%)

      Query: 7 VVYFSSKSNNNTHRFVQKLACSNQRIPSDGSSILVPTEDYIILIVPTYAGGGDDTKGAVPKQV 66
              +VYFSSKSNNNTHRFVQKL QRIP D + V+ Y+LIVPTYA GG D KGAV KQV
      Sbjct: 6 IVYFSSKSNNNTHRFVQKLGLPAQRIPVDNRPLEVSTHYLLIVPTYAAGGSDAKGAVSKQV 65

35      Query: 67 VQFLNVRQNREHCQGVISSGNNTNGDTYAIAGPIIARKLNVPPLLHQFELLGTQEDVTRVK 126
              ++FLN NR+HC+GVISSGNNTNGDT+A+AGPII++KL VPLLHQFELLGT DV +V+
      Sbjct: 66 IRFLNNPPNNRKHCKGVISSGNNTNGDTFALAGPIISQKLQVPLLHQFELLGTATDVKKVQ 125

40      Query: 127 ELLCQ 131
              + +
      Sbjct: 126 AIFAR 130

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 256

A DNA sequence (GBSx0271) was identified in *S.agalactiae* <SEQ ID 811> which encodes the amino acid sequence <SEQ ID 812>. This protein is predicted to be ribonucleoside-diphosphate reductase beta chain (nrdF). Analysis of this protein sequence reveals the following:

```

50 Possible site: 27
    >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3889 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-322-

>GP:AAB96162 GB:AE000050 ribonucleoside-diphosphate reductase beta
chain [Mycoplasma pneumoniae]
Identities = 261/335 (77%), Positives = 301/335 (88%)

5 Query: 2 QSYYDRSQSPLDYALSEKA~~FPMRSVNWNKLNNDDK~~DLEVWNRVTQNFWLPEKIPVSNDLNS 61
+ Y+ S SPL+YA + +RSVWN +D+KDLEVWNR+TQNFWLPEKIPVSND+ S
Sbjct: 5 KKYFLESVSPLEYAQKKFQGNLRSVNWNLV~~DDEK~~DLEVWNRITQNFWLPEKIPVSNDIPS 64

10 Query: 62 WRTLDADWQQLITRTFTGLTLLDSVQATVGDI~~AQIKHSQTDHEQVIYANFAFMVAIHARS~~ 121
W+ L +WQ LIT+TFTGLTLLD++QAT+GDI QI ++ TDHEQVIYANFAFMV +HARS
Sbjct: 65 WKQLSKEWQDLITKTFTGLTLLDTI~~QATIGDIKQIDYALT~~DHEQVIYANFAFMVGVHARS 124

15 Query: 122 YGTIFSTLCTSQQIEEAHEWWVVDTESLQARSRLI~~PFYTGDDPLKS~~KVAAAMMPGFLLYG 181
YGTIFSTLCTS+QI EAHEWWVV TESLQ R++ LIP+YTG DPLSKVAAA+MPGFLLYG
Sbjct: 125 YGTIFSTLCTSEQITEAHEWWVKTESLQ~~RAKALI~~PYYTGKDPLKS~~KVAA~~ALMPGFLLYG 184

20 Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKV~~IHNYYSGYKQQK~~VAKLSVEKQAEMKTFVFD 241
GFYLPFYLS+R +LPNTSDIIRLILRDKV~~IHNYYSGYK+Q+KV~~ K+S EKQAEMK FVFD
Sbjct: 185 GFYLPFYLSRKQLPNTSDIIRLILRDKV~~IHNYYSGYK~~FORKVEKMSKEKQAEMKRFVFD 244

25 Query: 242 LLYQLIDLEKAYLYELYDGF~~DAEDAIRFSIYNAGKF~~LQNLGYDSPFTEETRISPEVFA 301
L+Y+L+LEKAYL ELY+GF + EDAI+F~~SIYNAGKF~~LQNLGYDSPFTEETRI PE+FA
Sbjct: 245 LMYELIELEKAYL~~KELYEGFGIVEDAIKFSIYNAGKF~~LQNLGYDSPFTEETRIKPEIFA 304

30 Query: 302 QLSARADENH~~FFFSGNGSSYIMGITEETLDEDWEF~~ 336
QLSARADENH~~FFFSGNGSSY+MGI+EET~~ D+DW+F
Sbjct: 305 QLSARADENH~~FFFSGNGSSYVMGISEETEDKDWD~~ 339

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 813> which encodes the amino acid sequence <SEQ ID 814>. Analysis of this protein sequence reveals the following:

Possible site: 18
>> Seems to have no N-terminal signal sequence

----- Final Results -----

35 bacterial cytoplasm --- Certainty=0.3779(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

40 Identities = 292/335 (87%), Positives = 318/335 (94%)

Query: 2 QSYYDRSQSPLDYALSEKA~~FPMRSVNWNKLNNDDK~~DLEVWNRVTQNFWLPEKIPVSNDLNS 61
Q YY+RSQSP++YALSE +RS+NWN LNDDKDLEVWNRVTQNFWLPEK+PVSNDLNS
Sbjct: 3 QHYYERSQSP~~EYALSETQ~~QQLRSINWN~~YLNNDDK~~DLEVWNRVTQNFWLPEKVPVSNDLNS 62

45 Query: 62 WRTLDADWQQLITRTFTGLTLLDSVQATVGDI~~AQIKHSQTDHEQVIYANFAFMVAIHARS~~ 121
WR+L DWQQLITRT+TGLTLLD+VQATVGD+AQI+HSQTDHEQVIY NFAFMV IHARS
Sbjct: 63 WRSLGEDWQQLITRTYTG~~LDTVQATVGDVAQIQHSQTDHEQVIYTNFAFMVG~~IHARS 122

50 Query: 122 YGTIFSTLCTSQQIEEAHEWWVVDTESLQARSRLI~~PFYTGDDPLKS~~KVAAAMMPGFLLYG 181
YGTIFSTLC+S+QIEEAHEWWV T+SLQ R+R+LIP+YTGDDPLKS~~KVAA~~AMMPGFLLYG
Sbjct: 123 YGTIFSTLCSSEQIEEAHEWWV~~STQSLQDRARVLIPYYTGDDPLKS~~KVAAAMMPGFLLYG 182

55 Query: 182 GFYLPFYLSARGKLPNTSDIIRLILRDKV~~IHNYYSGYKQQK~~VAKLSVEKQAEMKTFVFD 241
GFYLPFYLSARGK+PNTSDIIRLILRDKV~~IHNYYSGYKQQKVA+LS~~ EKQAEMK FVFD
Sbjct: 183 GFYLPFYLSARGKMPNTSDIIRLILRDKV~~IHNYYSGYKQQKVARL~~SPEKQAEMKAFVFD 242

60 Query: 242 LLYQLIDLEKAYLYELYDGF~~DAEDAIRFSIYNAGKF~~LQNLGYDSPFTEETRISPEVFA 301
LLY+LIDLEKAYL ELY GFD~~DAEDAIRFS+YNAGKF~~LQNLGY+SPFT+EETR+SPEVFA
Sbjct: 243 LLYELIDLEKAYL~~RELYAGFD~~DAEDAIRFSLYNAGKF~~LQNLGY~~ESPFTDEETRV~~SPEVFA~~ 302

65 Query: 302 QLSARADENH~~FFFSGNGSSYIMGITEETLDEDWEF~~ 336
QLSARADENH~~FFFSGNGSSY+MGITEET~~ D+DW+F
Sbjct: 303 QLSARADENH~~FFFSGNGSSYVMGITEETTDDDW~~ 337

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 257

A DNA sequence (GBSx0272) was identified in *S.agalactiae* <SEQ ID 815> which encodes the amino acid sequence <SEQ ID 816>. This protein is predicted to be rhamnosyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.1741(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

15 A related GBS nucleic acid sequence <SEQ ID 9583> which encodes amino acid sequence <SEQ ID 9584> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA32090 GB:AB010970 rhamnosyltransferase [Streptococcus mutans]
  Identities = 104/309 (33%), Positives = 173/309 (55%), Gaps = 21/309 (6%)
20
Query: 11 QINICLATYNGQKYLRQQQLDSIIQQGYTDWICLIRDDGSTDDTVAIKEYVNRDSRFIFI 70
++NI ++TYNGQ+++ QQ+ SI +Q + +W LIRDDGS+D T II ++ D+R FI
Sbjct: 2 KVNLIMSTYNGQEFTAQIQSIQKQTFFENWNLLIRDDGSSDGTPKIIADFAKSDARIRFI 61
25
Query: 71 NSNDDRKLGSRSFYELVNYYKKADFYVFSQDDDVWKENRLERYLEEAEKFNQELPLLVYS 130
N++ G ++FY L+ Y+KAD+Y FSDQDDVV +LE L EK N ++PL+VY+
Sbjct: 62 NADKRENFGVIKNFYTLLKYEKADYYFFSDQDDVVLPQKLELTLASVEKENNQIPLMVYT 121
30
Query: 131 NWTSVDEKLTVL-----KEHNPATVIQEIQIAFNQINGMVIMMNHELAKLWE--YRQIG 181
+ T VD L VL + H+ T + E++ N + G +M+NH LAK W+ Y +
Sbjct: 122 DLTVVDRDLQVLHDSMIKTQSHANTSLLELTENTVTGGTMVMNHCLAKQWKQCYDDLI 181
35
Query: 182 AHDSYVGTLAYAVGNVAYISDSTVLWRRQ---VGAES----LNNYGRQYG-VATFWQMI 232
HD Y+ LA ++G + Y+ ++T L+R+ +GA + L N+ R + V +W ++
Sbjct: 182 MHDWYLALLAASLGKLIYLDETTELYRQHESNVLGARTWSKRLKNWLWPHRLVKKYWWLV 241
40
Query: 233 NTSFDRASLIFAQVSDKMSLERKLFFSRFIELKNANLMRRYLLSKLKLRRKSLKETVAM 292
+S +AS + + + K ++ L + + + RI L + + T
Sbjct: 242 TSSQQQASHL---LELDLPAANKAIIRAYVTLDDQSFLNRIKWLKQYGFAKNRAFHTFVF 298
45
Query: 293 TILLLTGYG 301
L++T +G
Sbjct: 299 KTLIITKFG 307
```

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 258

A DNA sequence (GBSx0273) was identified in *S.agalactiae* <SEQ ID 819> which encodes the amino acid sequence <SEQ ID 820>. Analysis of this protein sequence reveals the following:

```
50 Possible site: 36
>>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -4.19 Transmembrane 1213 -1229 (1211 -1230)

----- Final Results -----
55      bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
```

-324-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9581> which encodes amino acid sequence <SEQ ID 9582>
 5 was also identified.

There is also homology to SEQ ID 822.

A related GBS gene <SEQ ID 8525> and protein <SEQ ID 8526> were also identified. Analysis of this
 protein sequence reveals the following:

```

10    Lipop: Possible site: -1 Crend: 7
      SRCFLG: 0
      McG: Length of UR: 3
            Peak Value of UR: 2.28
            Net Charge of CR: 4
      15    McG: Discrim Score: 1.29
      GvH: Signal Score (-7.5): 2.84
            Possible site: 30
            >>> Seems to have a cleavable N-term signal seq.
            Amino Acid Composition: calculated from 31
            ALOM program count: 0 value: 1.16 threshold: 0.0
            20    PERIPHERAL Likelihood = 1.16      344
            modified ALOM score: -0.73

            *** Reasoning Step: 3

25    ----- Final Results -----
            bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
            bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

30 LPXTG motif: 1197-1201

SEQ ID 8526 (GBS147) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 extract is shown in Figure 29 (lane 4; MW 132kDa).

The GBS147-His fusion product was purified (Figure 200, lane 5) and used to immunise mice. The
 35 resulting antiserum was used for FACS (Figure 286), which confirmed that the protein is immunoaccessible
 on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 vaccines or diagnostics.

Example 259

40 A DNA sequence (GBSx0274) was identified in *S.agalactiae* <SEQ ID 823> which encodes the amino acid
 sequence <SEQ ID 824>. This protein is predicted to be Acetyltransferase (GNAT) family. Analysis of this
 protein sequence reveals the following:

```

45    Possible site: 57
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
            bacterial cytoplasm --- Certainty=0.2781 (Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
            bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG03505 GB:AE004449 conserved hypothetical protein [Pseudomonas aeruginosa]

-325-

Identities = 66/143 (46%), Positives = 94/143 (65%), Gaps = 5/143 (3%)

Query: 2 WNVKTFDNLTTHELFQIYKLRSVVFVVEQDCPYQEVDDEDLI--CLHGMNWVDGQLAAAYY 59
W K +LT EL+ +LR VFVVEQ CPYQEVD DL+ H M W DGQL AY

5 Sbjct: 5 WTCKHHADLTLKELYALLQLRTEVFVVEQKCPYQEVDGLDVLGDTHHLMAWRDGQLILAYL 64

Query: 60 RLIP---EDDKVHLGRVIVNPDFRKKGNGNQLVEYAIKFSEANYPNKPIYAQAQAYLQDF 116
RL+ + +V +GRV+ + R +GLG+QL+E A++ +E + + P+Y AQA+LQ +

10 Sbjct: 65 RLLDPVRHEGQVVIGRVVSSAARGQGLGHQILMERALQAAERLWLDTPVYLSAQAHQAY 124

Query: 117 YQSFGFQPVSIDIYLEDNIPHLD 139
Y +GF V+++YLED+IPH+ M

Sbjct: 125 YGRYGFVAVTEVYLEDHIPHIQM 147

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 260

A DNA sequence (GBSx0275) was identified in *S.agalactiae* <SEQ ID 825> which encodes the amino acid 20 sequence <SEQ ID 826>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2010 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

30 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 261

A DNA sequence (GBSx0276) was identified in *S.agalactiae* <SEQ ID 827> which encodes the amino acid 35 sequence <SEQ ID 828>. Analysis of this protein sequence reveals the following:

Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

40 bacterial cytoplasm --- Certainty=0.2935 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

45 >GP:CAB12631 GB:Z99108 similar to RNA methyltransferase [Bacillus subtilis]
Identities = 217/448 (48%), Positives = 298/448 (66%), Gaps = 4/448 (0%)

Query: 7 QRIPLKIKRMGINENGEGIFYKKTLIFVPGALKGEEVFCQISSVRRNFIAEKLLKINKKSK 66
Q PL IKR+GINGEG+G++KK ++FVPGAL GEEV Q + V+ F+E ++ KI K S+

50 Sbjct: 16 QTFFLTIKRLGINENGEGVGYFKKKVVFVPGALPGEVVVQATKVQPKSEGRIKKIRKASE 75

Query: 67 NRVEPPCSIYKECGGCQIMHLQYDKQLEFKTDVIRQALMKFKPEGYENYEIRKTIGMSEP 126
+RV PPC +Y++CGGCQ+ HL Y +QL K D++ Q+L + EN EI++TIGM P

-326-

Sbjct: 76 HRVAPPYVQECCGCQLQHLAYSQQLREKRDIVIQLSERHTKFV
VENMEIKETIGMDNP 135

Query: 127 EHYRAKLQFQV-RSFGGNVKAGLYAQGTHRLIDIKDCLVQDSL
TQEMINRVAELLGKYKL 185
+YR K QFQ+ RS G++ AGLY +H ++ IKDC+VQ T + V +L + +

5 Sbjct: 136 WNYRNKSQFQIGRSQSGSIIAGLYGLDSH DIVPIKDCIVQHPATNK
TTGIVRRILED FNV 195

Query: 186 PIYNERKIAG-VRTVMIRRAQASGEVQLIFITSKRL--DFDDVVIELVREFPELK
TVAVN 242
+YNERK G VRT++ R +GEVQ++ +T+K +++V ++ PE+K++ N

10 Sbjct: 196 SVYNERKRKGDVRTIVTRVGFETGEVQVVLVTAKETLPHKEEIVKAIQ
KRLPEVKSIIQN 255

Query: 243 INASKTS DIYGQITEVIWGQESINEEVLDYGFSLSPRAFYQLNP
KQTQILYSEAVKALDV 302
+N +KTS I+G+ T+ +G+ I E+ D F LS RAF+QLNP+QT LY E KA +

Sbjct: 256 VNGAKTSVIFGEGTKQLAGKTVIQEVLDVSFEL SARAFFQLNPE
QTVKLYDEVKKAQ 315

15 Query: 303 KEDDDLIDAYCGVGTIGLAGFKVSKVRGM DI IPEAIQD
AKENALYMGFTNTHYEAGKAE 362
+ ++DAYCGVGTIG+ A K VRGMD+I E+I DAK+NA G N Y G AE

Sbjct: 316 TGKEKVVDAYCGVGTIGMWADGAKEVRGMDVIKESIDDAKKH
GMANATYVTGTAE 375

20 Query: 363 DIIIPRWYSEGFRANALIVDPPTG LDDKLLN
TILKMPPEK MVYVSCNTSTLARDLVTLLTK 422
+P+W EGFR + IVDPPTG D L+TI K+ P++ VYVSCN STLA+DL TL+K

Sbjct: 376 HWLPKWTKEGFRPDVVIVDPPRTGCDSTFLDTIKKV
PKRFVYVSCNPSTLAKDLQTL SK 435

Query: 423 VYHVHYIQSVDMFPHTARTEAVVKLQRK 450

Y V YIQ VDMFP TA EAV +L K

25 Sbjct: 436 DYRVDYIQPVDMFPQTAHVEAVARLVLK 463

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 829> which encodes the amino acid sequence <SEQ ID 830>. Analysis of this protein sequence reveals the following:

Possible site: 56

30 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2980(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 327/450 (72%), Positives = 397/450 (87%)

40 Query: 1 MNVVLKQRIPLKIKRMGINGEGIGFYKKT
LIFVPGALKGEEVFCQISSVRRNFAEAKLLK 60
M V +KQ+ I PLKIKRMGINGEGIGFY+KTL+FVPGALKGE++FCQI++V+RNFAEAKLL
Sbjct: 1 MVVKVKQKIPPLKIKRMGINGEGIGFYQKTLV
FVPGALKGEDIFCQITAVKRNF A EAKLLT 60

45 Query: 61 INKKSKNRV
EPPCSIYKECGGCQIMHLQYDKQLEFKTDVIRQALMKFKPEGYENYEIRKT 120
+NK SKNRV+P CS+Y+ CGGCQIMHL Y QKL+FK DVIRQAL KFKP GYE +EIR T
Sbjct: 61 VNKASKNRVKPACSVYETCGGCQIMHLAYPKQLDFKDD
VIRQALKFKPTGYEQFEIRPT 120

50 Query: 121 IGMSEPEHYRAKLQFQVRSFGGNVKAGLYAQGTHRLIDIKD
CLVQDSL TQEMINRVAELL 180
+GM +P+HYRAKLQFQ+RSFGG VKAGL++QG+HRL+ I +CLVQD LTQ++IN++ +L+
Sbjct: 121 LGMKKPDHYRAKLQFQLRSFGGTVKAGLFSQGSHRLVPIDN
CLVQDOL TQDIINKITQLV 180

55 Query: 181 GKYKLPIYNERKIAGV
RTVMIRRAQASGEVQLIFITSKRLDFDDVVIELVREFPELK
VVA 240
KYKLPIYNERKIAG+RT+M+R+AQS +VQ+I ++SK + + + EL + FP++KTVA
Sbjct: 181 DKYKLPIYNERKIAGIRT
IMVRKAQASDQVIIIVSSKEVRLANFIGELTKAF
PQVKTV 240

Query: 241 VNINASKTS
DIYGQITEVIWGQESINEEVLDYGFSLSPRAFYQLNP
KQTQILYSEAVKAL 300
+N N SK+S+IYG TE++WGQE+I+EEVLDYGF+LSP
RAFYQLNP+QT++LY E VKAL
Sbjct: 241 LNSNRSKSSEIYGD
ETEILWGQEA
IHEEVLDYGFALSPRAFYQLNP
QQTEVLYGEVVKAL 300

60 Query: 301 DVKE
DDDLIDAYCGVGTIGLAGFKVSKVRGM
DI IPEAIQD
AKENALYMGFTNTHYEAGK 360
DV D +IDAYCGV+IG AFAGK
VSKVRGM DI IPEAI+DA++NA MGF N +YEAGK
Sbjct: 301 DVGS
KDHIIDAYCGVGSIGFA
FAGK
VSKVRGM DI IPEAI
EDAQNAKAMGF
DNAYYHAGK 360

65 Query: 361 AEDIIPRWYSEGFRANALIVDP
PTG LDDKLLN
TILKMPPEK
MVYVSCNTSTLARDL
V L
AEDI +WY +G+RA+A+IVDP
PTG LDDKLL TIL P++MVYVSCNTSTLARDL
V L

-327-

```

Sbjct: 361 AEDIISKWYKQGYRADAIVDPPRTGLDDKILLKTIHLHYQPQKQMVYVSCNTSTLARDLVQL 420
Query: 421 TKVYHVHYIQSVDMFPHTARTEAVVKLQRK 450
      TKVY VHYIQSVDMFPHTARTEAVVKLQRK++
5   Sbjct: 421 TKVYDVHYIQSVDMFPHTARTEAVVKLQRK 450

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 262

10 A DNA sequence (GBSx0277) was identified in *S.agalactiae* <SEQ ID 831> which encodes the amino acid sequence <SEQ ID 832>. Analysis of this protein sequence reveals the following:

```

Possible site: 45
>>> Seems to have no N-terminal signal sequence

```

```

15 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3505 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

20 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAB04643 GB:AP001510 unknown conserved protein in B. subtilis
      [Bacillus halodurans]
      Identities = 74/263 (28%), Positives = 141/263 (53%), Gaps = 9/263 (3%)

```

```

25 Query: 3 ITKIEKKKR---LYTLEL-DNTENLY---ITEDTIVHFMLSKGMIINAEKLENKKFAQL 55
      IT+IE +KR Y + + N +++Y + E ++ L KG+ I+AE+++ I ++
      Sbjct: 4 ITRIEVQKRNNERYNIFIHQNGQDVYAFSVDEQVLIKQGLRKGLDIDAEQMKQILYEDEV 63

```

```

30 Query: 56 SYGKNLGLYYISFKQRTEKEVIKYLQQHDIDSKIIIPQIIDNLKSENWINDKNYQSFITQQ 115
      NL L+Y+S++ R+ EV YL++ D + II ++ L + ++D + ++FIQ
      Sbjct: 64 QKTFNLALHYSYRMRSVHEVRTYLLKKKDREEPPIIEHVLHRLTEQRLDDHAFAAFIQT 123

```

```

35 Query: 116 NLNTGDKGPYVIKQKLLQKGIKSKIIIESLQAINFQDLASKISQKLYKKYQNKPLKAL- 174
      T KGP +KQ+L +KG+ K IE L +++++ ++ L K+ +L
      Sbjct: 124 KRATTSKGPLKLKQELAEGVSEKTIEGALTTFSYEEQVEQVKAWLEKQKGRTFKGSSLA 183

```

```

40 Query: 175 -DKKLMQSLLTKGFDYQIVHTVIQNLEIEKDQELEDLYIKELDKQYQKLSKKHDQYELK 233
      K KL + L KG+ ++ ++ I++++E E + + +K +K +K +EL+
      Sbjct: 184 WKQKLSRQLLAKGYTSPVIEEAFAVPIKQEEEEWEALKAFGEKAMRKYAGKKTGWELQ 243

```

```

Query: 234 QRIINALMRKGYQYEDIKSALRE 256
      Q++ AL RKG+ E I+ L +
      Sbjct: 244 QKVKQALYRKGFSELEMIERYLNND 266

```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 833> which encodes the amino acid sequence <SEQ ID 834>. Analysis of this protein sequence reveals the following:

```

Possible site: 58
>>> Seems to have no N-terminal signal sequence

```

```

50 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2388 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

55 An alignment of the GAS and GBS proteins is shown below:

```

Identities = 146/258 (56%), Positives = 190/258 (73%)

```

```

Query: 1 MKITKIEKKKRLYTLELDNTENLYITEDTIVHFMLSKGMIINAEKLENKKFAQLSYGKN 60
      MKITKIEKKKRLY +ELDN E+LY+TEDTIV FMLS K +++ ++LE++K FAQLSYGKN

```

5 Sbjct: 1 MKITKIEKKRKYLYLIEDNDESPLYVTEDTIVRFMLSKDKVLDNDQLEDMKHFAQLSYGKN 60
 Query: 61 LGLYYISFKQRTEKEVIKYLQQHDIDSKIIIPQIIDNLKSENWINDKNYVQSFIQQNLNTG 120
 L LY++SF+QR+ K+V YL++H+I+ II II L+ E WI+D ++I+QN G
 Sbjct: 61 LALYFLSFQQRSNKQVADYLRKHEIEEHIIADIITQLQEEWIDDTKLADTYIRQNQLNG 120
 10 Query: 121 DKGPYVIKQKLLQKGKIKSKIIIESELQAINFQDLASKISQKLYKKYQNLPLKALKDKLMQ 180
 DKGP V+KQKLLQKG I S I+ L +F LA K+SQKL+ KYQ KLP KALKDK+ Q
 Sbjct: 121 DKGPQVLQKQKLLQKGIA SHDIDPILSQTDFSQLAQKVSQQLFDKYQEKLPPKALKDKITQ 180
 15 Query: 181 SLTTKGFDYQIVHTVIQNLIEKDQELEEDLIYKELDKQYQKLSKKHDQYELKQRIIINAL 240
 +L TKGF Y + +L ++D + EDL+ KELDKQY+KLS+K+D Y LKQ++ AL
 Sbjct: 181 ALLTKGFSYDLAKHSLNHLDQDNQEIEDLLDKELDKQYRKLSRKYDGYTLKQKLYQAL 240

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 263

A DNA sequence (GBSx0278) was identified in *S.agalactiae* <SEQ ID 835> which encodes the amino acid sequence <SEQ ID 836>. Analysis of this protein sequence reveals the following:

25 Possible site: 33
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3912 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04659 GB:AP001510 unknown conserved protein in *B. subtilis*
 [Bacillus halodurans]
 35 Identities = 96/175 (54%), Positives = 122/175 (68%)
 Query: 1 MRLPKEGDFITIQSYKHDGSLHRTWRDTMVLKTENALIGVNDHTLVENDGRRWVTREP 60
 M PK G I IQSYKH+GS+HR W +T+VLIK T +IG ND LV E+DGR W TREP
 Sbjct: 1 MNFPKVGSKIQIQSYKHNGSIHRIWEETIVLKGTSKVVGNDRILVKESDGRHWRTREP 60
 40 Query: 61 AIVYFHKKYWFNIIAMIRETGVSYYCNLASPYILDPEALKYIDYDLDVKVFADGEKRLLD 120
 AI YF + WFN I MIR G+ +YCNL +P+ D EALKYIDYDLD+KVF D +LLD
 Sbjct: 61 AICYFDSEQWFTNTIGMIRADGIYFYCNLGPFTWDEEALKYIDYDLDIKVFPDMTFKLLD 120
 45 Query: 121 VDEYEQHKAAQMNPYPTDIDYILKENVKILVEWINENKGPFSSSYINIWYKRYLELK 175
 DEY H+ M YP +ID IL+ +V LV WI++ KGPF+ ++ WY+R+L+ +
 Sbjct: 121 EDEYAMHRKMMKYPPEIDRILQRSVDELVSWIHQRKGPFAPQFVESWYERFLQYR 175

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 837> which encodes the amino acid sequence <SEQ ID 838>. Analysis of this protein sequence reveals the following:

Possible site: 33
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 55 bacterial cytoplasm --- Certainty=0.3912 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 155/177 (87%), Positives = 165/177 (92%)

```

Query: 1 MRLPKEGDFITIQSQYKHD GSLHRTWRDTMVLKTTENALIGVNDHTLVTENDGRRWVTREP 60
      M+LPKEGDFITIQSQYKHD GSLHRTWRDTMVLKTTENALIGVNDHTLVTEDGRRWVTREP
5   Sbjct: 1 MKLPKEGDFITIQSQYKHD GSLHRTWRDTMVLKTTENALIGVNDHTLVTESDGRRWVTREP 60

Query: 61 AIVYFHKKYWFNIIAMI RETGVSYCNLASPYILDPEALKYIDYDLDVKVFADGEKRLLD 120
      AIVYFHKKYWFNIIAMI R+ GVSYYCNLASPY+D EALKYIDYDLDVKVFADGEKRLLD
10  Sbjct: 61 AIVYFHKKYWFNIIAMI RDNGVSYCNLASPYMDTEALKYIDYDLDVKVFADGEKRLLD 120

Query: 121 VDEYEQHKAQMNPYPTDIDYILKENVKILVEWINENKGPFSSSYINIWYKRYLEKKR 177
      VDEYE HK +M Y D+D+ILKENVKILV+WIN KGPFS +YI IWYKRYLELK R
      Sbjct: 121 VDEYEIHKKEMQYSADMDFILKENVKILVDWINHEKGPF SKAYITIWYKRYLELKNR 177

```

- 15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 264

A DNA sequence (GBSx0288) was identified in *S.agalactiae* <SEQ ID 839> which encodes the amino acid sequence <SEQ ID 840>. This protein is predicted to be jag protein. Analysis of this protein sequence 20 reveals the following:

```

Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
25 bacterial cytoplasm --- Certainty=0.1666 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

30 >GP:BAB07782 GB:AP001520 spoIIIJ-associated protein [Bacillus halodurans]
      Identities = 54/198 (27%), Positives = 98/198 (49%), Gaps = 6/198 (3%)

      Query: 100 DVVEEYIEEVDETLKEDVSQPELPKIDDKNVVTSEAIEKIDLLPNIEAAAQVTKYVE 159
              + VE+ I E+ T E+ + E PK ++ + A+ ++ + P+ + + +E
35   Sbjct: 13 EAVEQAIIEELGTTTRERITYTVVVEPKSGLFGILGSKP AVIEVVVKPD---PVDR AKAFL 69

      Query: 160 NIIYEMDLDA--TIETTSKRQINLQIETPEAGRIIGYHGKVLKSLQLLAQNYLHDRFSK 217
              ++ EMD++ TIE + N+ E + G +IG G+ L SLQ L + +
      Sbjct: 70 ELLQEMDMEVEV TIEKDPA TVL FNISGEO-DLGTLIGKRGQTLDLSLQYLVNLVANKEEG 128

40      Query: 218 SFSVSINVHDYVEHR TETLIDFSKKIARRVLETNEPYHMDPMNSERKTVHKTIA TIEGV 277
              + ++ +Y R E L+ + + +A + L T P + +PMS ERK +H + + V
      Sbjct: 129 FIRIKLDAENYRARRKEALVQLAERLASKALRTKRPV SLEPM SAHERKI IHTALQELGDV 188

45      Query: 278 ESYSEGNDPNRFVVVTKK 295
              E+YSEG R VV+ K
      Sbjct: 189 ETYSEGQQGIGRHHVIAPK 206

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 841> which encodes the amino acid 50 sequence <SEQ ID 842>. Analysis of this protein sequence reveals the following:

```

Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55 bacterial cytoplasm --- Certainty=0.3721 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

-330-

Identities = 176/302 (58%), Positives = 223/302 (73%), Gaps = 32/302 (10%)

Query: 23 MVLFTGATVEEAIKEKGLQELNISRLRAHIKVVSREKKGFLGFGKKPAKVEIEGITDEVTD 82
 MVLFTG TVEEAIE GLQEL +SRL+AHIKV+S+EKKGFLGFGKKPA+V+IEGI+D+

5 Sbjct: 1 MVLFTGKTVEEAIETGLQELGLSRLKAHIKVISKEKKGFLGFGKKPAQVDIEGISDKTVY 60

Query: 83 INESVALKNI-----KNVPS--SVDVVEEYIEEVDETLEKEDVSQPELPKIDDK---- 129
 + A + + +N P+ S DV E I+ + LE ED L D

10 Sbjct: 61 KADKKATRGVPEDINRQNTPAVNSADVEPEEKAT-QRLEAEDTKVVPLMSEDSPAQTPS 119

Query: 130 ---NVVTTSEA-----IEKIDL-----LPNIEAAAQVTKYVENIIYEMDLDATI 171
 VT ++A +E+ ++ +IE AA +V+ YV IIYEMD++AT+

Sbjct: 120 NLAFTVTETKAQQPSIPVEESEVPQDAGNDGFSKDIKAAQEVSVDYVTKIIYEMDIEATV 179

15 Query: 172 ETTTSKRQINLQIETPEAGRRIIGYHGKVLKSLQLLAQNYLHDQRSKSFSVSINVHDYVEH 231
 ET+ ++RQINLQIETPEAGR+IGYHGKVLKSLQLLAQN+LHDR+SK+FSVS+NVHDYVEH

Sbjct: 180 ETSNNRRQINLQIETPEAGRIVIGYHGKVLKSLQLLAQNFHLDQRSKNFSVSINVHDYVEH 239

20 Query: 232 RTETLIDFSKKIARRVLETNEPYHMDPMSNSERKTVHKTIATIEGVESYSEGNDPNRFVV 291
 RTETLIDF++K+A+RVLE+ + Y MDPMNSERK VHKT+++IEGV+SYSEGNDPNR+VV

Sbjct: 240 RTETLIDFTQKVAKRVLESQDYTMDPMSNSERKIVHKTVSSIEGVDSYSEGNDPNRYVV 299

Query: 292 VT 293

V+

25 Sbjct: 300 VS 301

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 265

30 A DNA sequence (GBSx0290) was identified in *S.agalactiae* <SEQ ID 843> which encodes the amino acid sequence <SEQ ID 844>. This protein is predicted to be 60 kd inner-membrane protein (yidC). Analysis of this protein sequence reveals the following:

Possible site: 42

>>> May be a lipoprotein

35 INTEGRAL Likelihood = -7.38 Transmembrane 54 - 70 (52 - 75)
 INTEGRAL Likelihood = -5.20 Transmembrane 193 - 209 (192 - 211)
 INTEGRAL Likelihood = -3.61 Transmembrane 125 - 141 (124 - 144)
 INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 (167 - 184)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.3951(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA78595 GB:Z14225 SpoIIJJ [Bacillus subtilis]

Identities = 79/243 (32%), Positives = 142/243 (57%), Gaps = 5/243 (2%)

Query: 1 MKKKLKTFSLLLTGSLLVACG--RGEVSSHSATLWEQ-IVYAFAKSIQWLS--FNHSIG 55
 MK+++ ++ LL C + + + S W++ +VY ++ I + + + + G

50 Sbjct: 1 MKRRIGLLMSMVGVFMALLAGCSSVKEPITADSPHFWDKYVVYPLSELITYVAKLTGDNYG 60

Query: 56 LGTILFTLITIRAIMMPLYNMQMKSQKMQEIOPRLKELQKKYPGKDPIDNRKLNLDEMQSM 115
 L IIL T++IR +++PL Q++SS+ MQ +QP +++L++KY KD + KL E ++

55 Sbjct: 61 LSTILVTLIRLLILPLMTKQLRSSKAMQALQPEMQLKEKYSSKDQKTQQKLOQETMAL 120

Query: 116 YKAEGVNPNPYASVPLLIQLPVILWALFQALTRVSFLKVGTFLSLELSQDPYYILPVLAAL 175
 ++ GVNP A P+LIQ+P+L + A+ R + +FL +L + DPYYILP++A +

Sbjct: 121 FQKHGVNPLAGCFPILIQMPLIGFYHAIMRTQAISEHSFLWFDLGEKDYPYYILPIAVGV 180

60 Query: 176 FTFLSTWLTNKAAVEKNIALTLMTYVMPFIILVTSFNFASGVVLYWTVSNAFQVFQILL 235

-331-

TF+ L ++N + +M ++MP +I+V + NF + + LYW V N F + Q L+
 Sbjct: 181 ATFVQQKLMAGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWWVGNLFMIAQTLI 240

5 Query: 236 NNP 238
 P
 Sbjct: 241 KGP 243

A related GBS sequence was identified <SEQ ID 10783> which encodes amino acid sequence <SEQ ID 10784>.

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 845> which encodes the amino acid sequence <SEQ ID 846>. Analysis of this protein sequence reveals the following:

Possible site: 49

15 >>> May be a lipoprotein
 INTEGRAL Likelihood = -6.32 Transmembrane 198 - 214 (197 - 220)
 INTEGRAL Likelihood = -5.52 Transmembrane 59 - 75 (57 - 80)
 INTEGRAL Likelihood = -4.25 Transmembrane 130 - 146 (129 - 150)
 INTEGRAL Likelihood = -2.28 Transmembrane 173 - 189 (170 - 189)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.3527 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:BAA05234 GB:D26185 stage III sporulation [Bacillus subtilis]
 Identities = 90/249 (36%), Positives = 150/249 (60%), Gaps = 6/249 (2%)

30 Query: 16 IVPLVLLLVACG--RGEVTAQSSSGWDQ-LVYLFRARAIQWL--FDGSIGVGIILFTLTI 70
 +V + +LL C + +TA S WD+ +VY + I +++ + G+ IIL T+ I
 Sbjct: 13 MVGVFMLLAGCSSVKEPITADSPHFWDKYVVYPLSELITYVAKLTGDNYGLSIIILVTILI 72

Query: 71 RLMMLPLFNMQIKSSQKMQDIQPELRELQRKYAGKDTQTRMKLAAEESQALYKKYGVNPYA 130
 RL+++PL Q++SS+ MQ +QPE+++L+ KY+ KD +T+ KL +E+ AL++K+GVNP A
 35 Sbjct: 73 RLLLILPLMIKQLRSSKAMQALQPEMQLKEKYSSKDQKQQKLQQETMALFQKHGVNPLA 132

Query: 131 SLLPLLIQMPVMIALFQALTRVSFLKTGTFLWELAQHDHLYLLPVLAATVFTFLSTWLTN 190
 P+LIQMP++I + A+ R + +FLW +L + D Y+LP++A V TF+ L
 Sbjct: 133 GCFPILIQMPILIGFYHAIMRTQAISEHSFLWFLGEKDPYYILPIVAGVATFVQQKLM 192

40 Query: 191 LAAKEKNVMMTVMIYVMPMLIFFMGFNLASGVVLYWTVSNAFQVVQLLLLNNP-FKIIAE 249
 +-N M +M++MP+MI N + + LYW V N F + Q L+ P K E
 Sbjct: 193 AGNAQQNPQMAMMLWIMPIMIIVFAINFPAALSLYWWVGNLFMIAQTLI 252

45 Query: 250 RQRLANEK 258
 Q+ ++K
 Sbjct: 253 PQKAGGKKK 261

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 172/270 (63%), Positives = 217/270 (79%), Gaps = 1/270 (0%)

Query: 1 MKKKLKTFSLILLTGSLLVACGRGEVSSHSATLWEQIVYAFAKSIQWL SFNHSIGLGIIL 60
 +KK +K ++ L LLVACGRGEV++ S++ W+Q+VY FA++IQWL SF+ SIG+GIIL
 Sbjct: 7 VKKNIKIARIIVPLV-LLL VACGRGEVTAQSSSGWDQLVYLFARAIQWL SFDGSIGVGIIL 65

55 Query: 61 FTLIIRAIMMPLYNMQMKSQKMQEIQPRIKELQKKYPGKDPDNRLKLNDDEM QSMYKAEG 120
 FTL IR ++MPL+NMQ+KSSQKMO+IQP L+ELQ+KY GKD R+KL +E Q++YK G
 Sbjct: 66 FTLLTIRLMLMPFLNMQIKSSQKMQDIQPELRELQRKYAGKDTQTRMKLAAEESQALYKKY 125

60 Query: 121 VNPyASVPLPLLQLPVLWALFQALTRVSFLKVGTFLSLELSQPDPPYYILPVLAALFTFLS 180
 VNPyAS+LPLL IQ+PV+ ALFQALTRVSFLK GTFL +EL+Q D Y+LPVLA+FTFLS
 Sbjct: 126 VNPyASLPLL IQMPVMIALFQALTRVSFLKTGTFLWELAQHDHLYLLPVLAAVFTFLS 185

Query: 181 TWLTNKAIVEKNIALTLMTYVMPFIILVTSFNFASGVLYWTVSNAFQVFQI₁LLNNPYK 240
 TWLTN AA EKN+ +T+M YVMP +I FN ASGVVLYWTVSNAFQV Q+LLLNNP+K
 Sbjct: 186 TWLTNLAAKEKNVMMTVM₁YVMPPLIFFMGFNLASGVVLYWTVSNAFQVVQ₁LLNNPK 245

5
 Query: 241 IIKVREEAVRVAHEKEQRVKRAKRKASKKR 270
 II R+ E+ R +RA++KA K++
 Sbjct: 246 II₁AERQR₂L₃NEEK₄R₅R₆R₇R₈R₉KAMKRK 275

- 10 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8527> and protein <SEQ ID 8528> were also identified. Analysis of this protein sequence reveals the following:

15 Lipop: Possible site: 20 Crend: 5
 McG: Discrim Score: 4.90
 GvH: Signal Score (-7.5): -0.39
 Possible site: 42
 >>> May be a lipoprotein
 20 ALOM program count: 4 value: -7.38 threshold: 0.0
 INTEGRAL Likelihood = -7.38 Transmembrane 54 - 70 (52 - 75)
 INTEGRAL Likelihood = -5.20 Transmembrane 193 - 209 (192 - 211)
 INTEGRAL Likelihood = -3.61 Transmembrane 125 - 141 (124 - 144)
 INTEGRAL Likelihood = -2.44 Transmembrane 168 - 184 (167 - 184)
 25 PERIPHERAL Likelihood = 2.54 217
 modified ALOM score: 1.98

*** Reasoning Step: 3

----- Final Results -----

30 bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 32.8/62.3% over 242aa
 Bacillus subtilis
 EGAD|17722| stage III sporulation protein j precursor Insert characterized
 OMNI|NT01BS4782 -identity Insert characterized
 SP|Q01625|SP3J_BACSU STAGE III SPORULATION PROTEIN J PRECURSOR. Edit characterized
 40 GP|40023|emb|CAA44401.1||X62539 unnamed protein product Insert characterized
 GP|467388|dbj|BAA05234.1||D26185 stage III sporulation Insert characterized
 GP|2636651|emb|CAB16141.1||Z99124 alternate gene name: spo0J87 Insert characterized
 PIR|I40437|I40437 stage III sporulation protein spoIIIJ - Insert characterized
 45 ORF02221(301 - 1014 of 1413)
 EGAD|17722| S4098(3 - 245 of 261) stage III sporulation protein j precursor { acillus
 subtilis}OMNI|NT01 S4782 -identity SP|Q01625|SP3J_ ACSU STAGE III SPORULATION PROTEIN J
 PRECURSOR.GP|40023|emb|CAA44401.1||X62539 unnamed protein product { acillus
 subtilis}GP|467388|dbj| AA05234.1||D26185 stage III sporulation { acillus
 subtilis}GP|2636651|emb|CA 16141.1||Z99124 alternate gene name: spo0J87 { acillus
 subtilis}PIR|I40437|I40437 stage III sporulation protein spoIIIJ - acillus subtilis
 %Match = 17.0
 %Identity = 32.8 %Similarity = 62.2
 Matches = 79 Mismatches = 88 Conservative Sub.s = 71
 55
 219 249 279 309 339 393 420
 DFVVIARKGVEELDYQALEKNL₁HVLKIAGLI*KG₂IKLKKKL₃TFS₄L₅LLTG₆SLLVACG--RGEVSSHSATLWEQ-IVYA
 : ||:: : :: || | : :: | :||: :: |||
 MLLKRRIG₁LLLSMVG₂FMLLAGC₃SSV₄KEPITADSPHFWDKYVVYP
 60
 10 20 30 40
 474 504 534 564 594 624 654
 FAKSIQWLS--FNHSIGLGIILFTLII₁RAIMMPLYNMQMKSSQKMQEIQ₂PR₃L₄KQKKY₅PGKD₆PD₇NRL₈KLNDEM₉QSMYKA

-333-

::|:: :|||::|::||| :|||::|::||| :|||::|::||| :|||::|::||| :|||::|::|||
 LSELITYVAKLTGDNYGLSIIILVTILIRLLILPIMIKQLRSSKAMQALQPEMQLKEKYSSKDQKTQQQLQOETMALFQK
 60 70 80 90 100 110 120
 684 714 744 774 804 834 864 894
 EGVNPYASVPLPLLQQLPVLWALFQALTRVSFLKVGTFLSLELSQPDPYYILPVLAALFTFLSTWLTNKAIVEKNIALTLM
 ||||| :|||:|||:||| :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 HGVNPLAGCFPIIQLOMPILIGFYHAIMRTQAISEHSFLWFDLGEKDYPYILPIVAGVATFVQOKLMMAGNAQONPQMAMM
 140 150 160 170 180 190 200
 924 954 984 1014 1044 1074 1104 1134
 TYVMPFIILVTSFNFAASGVVLYWTVSNAFQVFQIILLNNPYKIIKVREEAVRAHEKEQRVKRAKRKASKKRK*ENHGTT
 ::||:|||:|||:|||:|||:|||:|||:|||:|||:
 LWIMPIMIIVFAINFPAALSPLYWVGNLFMIAQTFLIKGPDIKKNPEPKAGGGKK
 220 230 240 250 260

Example 266

A DNA sequence (GBSx0291) was identified in *S.agalactiae* <SEQ ID 847> which encodes the amino acid sequence <SEQ ID 848>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3778(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9579> which encodes amino acid sequence <SEQ ID 9580> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA44400 GB:X62539 homologous to E.coli rnpA [Bacillus subtilis]
Identities = 52/109 (47%), Positives = 77/109 (69%), Gaps = 1/109 (0%)

Query: 21 LKKTYRKVKSDKDFQMFISRGKVNANRKFVIYYLEK-EQKHFRVGISVSKKLGNAVVRNAI 79
LKK R+K ++DFQ +F G +VANR+FV+Y L++ E RVG+SVSKK+GNAV+RN I
Sbjct: 4 LKKRNRLKKNEDFQKVFKHGTSVANRQFVLYTLDQPENDELRVGLSVSKKIGNAVMRNRI 63

Query: 80 KRKRIRHVLLSQKTAHQDVFVVIARKGVEELDYQALEKNLIHVLCIAGL 128
KR IR L +K L++ D+++IARK +L Y+ +K+L H+ + + L
Sbjct: 64 KRLIRQAFLLEKERLKEKDYIIIARKPASQLTYEETKKSLQHLFRKSSL 112

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 849> which encodes the amino acid sequence <SEQ ID 850>. Analysis of this protein sequence reveals the following:

Possible site: 24

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3820(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 73/109 (66%), Positives = 88/109 (79%)

Query: 21 LKKTYRKVKSDKDFQMIIFSRGKVNANRKFVIYYLEKEQKHFRVGIVSVKLGNNAVRRAIK 80
LKKTYRVK +KDFQ IF GK+ ANRKFVIY+L + Q HFRVGIVS KK+GNAV RNA+K
Sbjct: 1 LKKTYRVKREKDFQAIKDGSANRKFVIYHILNRGQDHFRVGIVSGKKIGNAVTRNAVK 60

Query: 81 RKIRHVLLSQKTAQDQDFVVIARKGVEELDYQALEKNLIHVHLKIAGLI 129
RKIRHV++ I+ DFWVIAKGV I+YO I+NI HVLK+A I+

-334-

Sbjct: 61 RKIRHVIMALGHQLKSEDFVVIARKGVHSLEYQELQQNLHHVLKLAQLL 109

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 267

A DNA sequence (GBSx0292) was identified in *S.agalactiae* <SEQ ID 851> which encodes the amino acid sequence <SEQ ID 852>. This protein is predicted to be glycerol-3-phosphate dehydrogenase, NAD-dependent (gpsA). Analysis of this protein sequence reveals the following:

Possible site: 33
 10 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1429 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 15 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8529> which encodes amino acid sequence <SEQ ID 8530> was also identified. There is a signal peptide at residues 1-19. The protein has homology with the following sequences in the GENPEPT database:

20 >GP:AAA86746 GB:U32164 NAD(P)H-dependent dihydroxyacetone-phosphate
 reductase [Bacillus subtilis]
 Identities = 177/333 (53%), Positives = 241/333 (72%)
 25 Query: 18 QKIAVLGPGSGWTALAQLNNDNGHEVRLWGNVVEQIEEINTNHTNQRYFKDITLDSKIKA 77
 +K+ +LG GSWGTALA VL DNG+EV +W + + I +IN H N+ Y ++ L + IK
 Sbjct: 2 KKVTMLGAGSGWTALALVLTNDNGNEVCVWAHADLIHQINELHENKDYLPNVKLSTSIIKG 61
 Query: 78 YTNNLEAINNVDSILFVVPTKVTRLVAKQVANLLKHKVVLMHASKGLEPGTHERLSTILE 137
 T+++EA+++ D I+ VPTK R V +Q + K V +H SKG+EP + R+S I+E
 30 Sbjct: 62 TTDMKEAVSDADVIIIVAVPTKAIREVLRQAVPFITKKAVFVHSKGIEPDSSLRISEIME 121
 Query: 138 EEISEQYRSDIVVVSGPSHAEEAIVRDITLITAASKDIEAAKYVQKLFNSHYFRLYTNTD 197
 E+ R DIVV+SGPSHAEE +R T +TA+SK + AA+ VQ LF NH FR+YTIN D
 Sbjct: 122 IELPSDVRRDIVVLSGPSHAEEVGLRHATTVTASSKSMRAAEVQDLFINHNFRVYTNPD 181
 35 Query: 198 VVGVETAGALKNIIAVGAGALIHLGYCDNAKAAITRGLAEITRLGVQLGADPLTFSGLS 257
 ++GVE GALKNIIA+ AG GLGYCDNAKAA+ITRGLAEI RLG ++G +PLTFSGL+
 Sbjct: 182 IIGVEIGGALKNIIALAAGITDGLGYCDNAKAALITRGLAEIARLGTKMGGNPLTFSGLT 241
 40 Query: 258 GVGDLIVTGTSVHSRNWRAGDALGRGEKLEDIEKNMGMVIEGISTTKVAYEIAQNLNVYM 317
 GVGDLIVT TSVHSRNWRAG+ LG+G KLED+ + MGMV+EG+ TTK AY++++ +V M
 Sbjct: 242 GVGDLIVTCTSVHSRNWRAGNLLGKGYKLEDVLEEMGMVVEGVRTTKAAYQLSKKDYDKM 301
 Query: 318 PITEAIYKSIVYEGANIKDSILDMMMSNEFRSENE 350
 PITEA+++ ++ G ++ ++ +M+ E E
 45 Sbjct: 302 PITEALHQVLFNQKQVETAVESLMARGKTHEME 334

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 853> which encodes the amino acid sequence <SEQ ID 854>. Analysis of this protein sequence reveals the following:

50 Possible site: 19
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0882 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 55 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

-335-

Identities = 287/338 (84%), Positives = 316/338 (92%)

Query: 15 MTKQKIAVLGPGSGWTALAQLAQVNLNDNGHEVRLWGNVVEQIEEINTNHTNQRYFKDITLDSK 74
MTKQK+A+LGPMSGWTAL+QVNLNDNGH+VRLWGN+ +QIEEINT HTN+ YFKDI LD

5 Sbjct: 1 MTKQKVALILGPGSGWTALSQVNLNDNGHDVRLWGNIPDQIEEINTKHTNRHYFKDIVLDKN 60

Query: 75 IKAYTNLEEAINNVDISLFVVPTKVTRLVAKQVANLLKHKVVLMHASKGLEPGTHERLST 134
I A +L +A+++VD+++LFVVPTKVTRLVA+QVA +L HKVV+MHASKGLEP THERLST

10 Sbjct: 61 ITATLDLGQALSDVDAVLFVVPTKVTRLVARQVAAILDHKVVMHASKGLEPETHRLST 120

Query: 135 ILEEEISEQYRS DIVVVS GPSHAA EAI VRDITLITAASKDIEAAKYVQKL FSNHYFRLYT 194
ILEEEI +RS++ +VVVSGPSHAA E IVRDITLITAASKDIEAAKYVQ LFSNHYFRLYT

Sbjct: 121 ILEEEI PAHFRSE VVVVSGPSHAEETIVRDITLITAASKDIEAAKYVQSLFSNHYFRLYT 180

15 Query: 195 NTDVVG VETAGALKNI IAVGAGALHGLGYGDNAAKAA IT RGLAEITRLGVQLGADPLTFS 254
NTDV+G VETAGALKNI IAVGAGALHGLGYGDNAAKAA+IT RGLAEITRLGV+LGADPLT+S

Sbjct: 181 NTDVIG VETAGALKNI IAVGAGALHGLGYGDNAAVITRGLAEITRLGVKLGADPLTYS 240

20 Query: 255 GLSGVGDLIVTGT SVHSRNWRAG DALGRGEKLEDIE K N GMVIEGISTTKVAYEIAQNLN 314
GLSGVGDLIVTGT SVHSRNWRAG ALGRGEKLEDIE +N GMVIEGI +TT KVAYEIAQ+L

Sbjct: 241 GLSGVGDLIVTGT SVHSRNWRAG ALGRGEKLEDIER N GMVIEGI ATT KVAYEIAQDLG 300

Query: 315 VYMPITEAIYKSIYEGANIKDSI LD MMSNEFRSENEWH 352

VYMPIT AIYKSIYEGA+IK+SIL MMSNEFRSENEWH

25 Sbjct: 301 VYMPITTAIYKSIYEGADIKESTI LGMMSNEFRSENEWH 338

SEQ ID 8530 (GBS291) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 59 (lane 5; MW 38.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 2; MW 64kDa).

30 GBS291-GST was purified as shown in Figure 226, lane 10-11.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 268

A DNA sequence (GBSx0293) was identified in *S.agalactiae* <SEQ ID 855> which encodes the amino acid 35 sequence <SEQ ID 856>. This protein is predicted to be glucose-1-phosphate uridylyltransferase (gtaB). Analysis of this protein sequence reveals the following:

Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

40 ----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA28714 GB:AB001562 glucose-1-phosphate uridylyltransferase
[Streptococcus mutans]
Identities = 263/296 (88%), Positives = 285/296 (95%)

50 Query: 2 KVRKAVIPAAGLGRFLPATKALAKEMLPIVDKPTIQFIVEALKSGIEDILVVTGKSKR 61
KVRKAVIPAAGLGRFLPATKALAKEMLPIVDKPTIQFIVEALKSGIEDILVVTGKSKR
Sbjct: 5 KVRKAVIPAAGLGRFLPATKALAKEMLPIVDKPTIQFIVEALKSGIEDILVVTGKSKR 64

55 Query: 62 SIEDHFDSNFELEYNLKEKGKNELLKLVDETTGIRLHFIRQSHPRGLGDAVLQAKAFVGN 121
SIEDHFDSNFELEYNL++KGK +LLKLV++TT I LHFIRQSHPRGLGDAVLQAKAFVGN
Sbjct: 65 SIEDHFDSNFELEYNLQKGKTDLLKLVNDTTAINLHFIRQSHPRGLGDAVLQAKAFVGN 124

Query: 122 EPFVVMLGDDLMITNNKVIPLTKQLINDFEATHASTIAVMEVPHEDVSAYGVIAPQGEG 181

-336-

EPFVVMLGDDLMIT++K IPLT+QL+ND+E THASTIAVMEVPHEDVSAYGVIAPQGEG
 Sbjct: 125 EPFVVMLGDDLMITDDKAIPLTRLQMLNDYEETHASTIAVMEVPHEDVSAYGVIAPQGEG 184

5 Query: 182 VNGLYSVNTFVEKPSPEEAPSNLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241
 V+GLYSV+TFVEKP+P+EAPSNLAIIGRYLLTPEIF ILETQ+PGAGNE+QLTDAIDTLN
 Sbjct: 185 VSGLYSVDTFVEKPKAPKEAPSNLAIIGRYLLTPEIFTILETQEPGAGNEVQLTDAIDTLN 244

Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKYYIIDLGKSLEKTS 297
 KTQRVFAR+F G RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII+LGK L++ S

10 Sbjct: 245 KTQRVFAREFKGKRYDVGDKFGFMKTSIDYALKHPQVKEDLKAYIIELGKKLDQKS 300

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 857> which encodes the amino acid sequence <SEQ ID 858>. Analysis of this protein sequence reveals the following:

Possible site: 26

15 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>

bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 257/295 (87%), Positives = 277/295 (93%)

25 Query: 2 KVRKAVIPAAGLGRFLPATKALAKEMLPIVDKPTIQFIVEALKSGIEDILVVTGSKR 61
 KVRKA+IPAAGLGRFLPATKALAKEMLPIVDKPTIQFIVEALKSGIE+ILVVTGK+KR

Sbjct: 3 KVRKAIIPAAGLGRFLPATKALAKEMLPIVDKPTIQFIVEALKSGIEEILVVTGAKR 62

30 Query: 62 SIEDHFDSNFELEYNLKEKGKNELLKLVDETTGIRLHFIRQSHPRGLGDAVLQAKAFVGN 121
 SIEDHFDSNFELEYNL+ KGKNEELLKLVDETT I LHFIRQSHPRGLGDAVLQAKAFVGN

Sbjct: 63 SIEDHFDSNFELEYNLQAKGKNELLKLVDETTAIDLHFIRQSHPRGLGDAVLQAKAFVGN 122

Query: 122 EPFVVMLGDDLMITNNKVIPLTKQLINDFEATHASTIAVMEVPHEDVSAYGVIAPQGEG 181
 EPFVVMLGDDLMITN PLTKQL+ D++ THASTIAVM+VPHEDVS+YGVIAAPQG+
 35 Sbjct: 123 EPFVVMLGDDLMITNASAKPLTKQLMEDYDKTHASTIAVMKVPHEVDSSYGVIAAPQGKA 182

Query: 182 VNGLYSVNTFVEKPSPEEAPSNLAIIGRYLLTPEIFNILETQKPGAGNEIQLTDAIDTLN 241
 V GLYSV+TFVEKP PE+APS+LAIIGRYLLTPEIF ILE Q PGAGNE+QLTDAIDTLN
 Sbjct: 183 VKGLYSVDTFVEKPKQPEDAPSDLAIIGRYLLTPEIFGILERQTPGAGNEVQLTDAIDTLN 242

40 Query: 242 KTQRVFARKFTGDRYDVGDKFGFMKTSIDYALQHPQVKDDLKYYIIDLGKSLEKTS 296
 KTQRVFAR+F G+RYDVGDKFGFMKTSIDYAL+HPQVK+DLK YII LGK+LEK+
 Sbjct: 243 KTQRVFAREFKGKRYDVGDKFGFMKTSIDYALEHPQVKEDLKNYIILGKALEKS 297

45 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 269

A DNA sequence (GBSx0294) was identified in *S.agalactiae* <SEQ ID 859> which encodes the amino acid sequence <SEQ ID 860>. Analysis of this protein sequence reveals the following:

50 Possible site: 42

>>> Seems to have no N-terminal signal sequence

INTEGRAL Likelihood = -4.94 Transmembrane 28 - 44 (27 - 45)

----- Final Results -----

55 bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

%Match = 21.7
 %Identity = 52.7 %Similarity = 73.8
 Matches = 125 Mismatches = 59 Conservative Sub.s = 50

5	294	324	354	384	414	444	465	489
	NFLWEK*NKVC*MIFLCYDRNLFLCDYNLLGGGSFSVNRKIIIGLTLLSLSVLTLACGNRSD---KSANKS--DIKVAMVT							
	: ::: : : :							
	MKKRTFALALSMIIASGVILGACGSSSSDKKSSDDKSSKDFTVAMVT							
	10 20 30 40							
10	519	549	579	606	636	666	696	726
	NQGGVDDDKSFNQSAWEGLQKGKKGLTKG-NGFDYFQSSNESDHANNLDTAASSGYNLIFGIGFGLHDTIEKVSEENKD							
	: : : : : : : : : : : : : : : : : :							
	DTGGVDDDRSFNQSAWEGLQKGKANDMEKGTDGYNYLQSASEADYKTNLNTAVRSODYDLIYGIGYKLKDAIEEVSKQPK							
	15 60 70 80 90 100 110 120							
20	756	786	816	846	876	906	936	966
	VKYVIVDDTIKGKENVASVTFADNEAAAYLAGVAAKTTKTKTIVGFIGGMEGVVVKRFEAGFKAGVKSIDPAIKVAVSYAG							
	: : : : : : : : ::							
	NQFAIVDDTIDDRDNVVSIGFKNDGSYLVGVVAGLTTKTNKVGFGVGGVKGTVIDRFEAGFTAGVKAVNPNAQIDVQYAN							
	140 150 160 170 180 190 200							
25	996	1026	1056	1086	1116	1146	1176	1206
	SFTDAAKGKTIAATQYATGVDVIYQAAGGTGAGIFSEAKTENETRKESNKVWWIGVDRDQSQEGNYVSKDGKKANFVLAS							
	: :							

A related GBS nucleic acid sequence <SEQ ID 10947> which encodes amino acid sequence <SEQ ID 10948> was also identified.

SEQ ID 8532 (GBS108) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 38 (lane 7; MW 39.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 41 (lane 9; MW 64.6kDa).

The GBS108-GST fusion product was purified (Figure 202, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 273), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 270

40 A DNA sequence (GBSx0295) was identified in *S.agalactiae* <SEQ ID 863> which encodes the amino acid sequence <SEQ ID 864>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have a cleavable N-term signal seq.
  INTEGRAL Likelihood = -12.74 Transmembrane 206 - 222 ( 197 - 224)
  INTEGRAL Likelihood = -3.72 Transmembrane 174 - 190 ( 171 - 194)
  INTEGRAL Likelihood = -3.19 Transmembrane 98 - 114 ( 98 - 116)
  INTEGRAL Likelihood = -1.54 Transmembrane 120 - 136 ( 120 - 139)
  INTEGRAL Likelihood = -0.90 Transmembrane 157 - 173 ( 157 - 173)
```

50 ----- Final Results -----
 bacterial membrane --- Certainty=0.6095 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
 Identities = 126/218 (57%), Positives = 166/218 (75%)

Query: 8 KEYPTTIVLLVSLTTLVFLLMQLTYGSQAESSQVIFQFGGIQQGDYLKAYPTNLWRLISPIF 67
 KE P T +S+T L+F++MQ+ YGS A+S QV+FQFGG+ G +K+ P+ LWRL++PIF
 Sbjct: 5 KEKPVTFFFSLSVTILLFIVMQLFYGSWAKSPQVVFQFGGMFLVVKSMPSQLWRLVTPIF 64

Query: 68 VHIGWEHFLNLALYFVGQMGESIWGSRLFLILYILSGLMGNIFTLFFTPHVVAAGAST 127
 +HIGWEHFL+N L LYFVGQ+ ESIWGS FL+LY+LSG+MGN+ TLFFTPHVVAAGAST

Sbjct: 65 IHIGWEHFLINSLTLYFVGQIAESIWSRFLLLVLSGIMGNVTLFFTPHVVAAGAST 124

Query: 128 SLFGVFSAIIAIAGYFGKNPYLKQVGKSYQVMILLNLFFNIFTPGVSLAGHVGGLV 187
 SLFG+F+AI + GYFG N LK +GKSYQ +I+LNL N+F P V + GH+GG +GG L

Sbjct: 125 SLFGLFAAIVVVGYFGHNQLLKSIGKSYQTIIILNLVMNLFMPNVGIVGHGALGGALA 184

Query: 188 AIFLTQNGSLLFKTWOSILALMIFIIVSISLIGLSLV 225
 A+FL + LF Q AL+ ++ +++ LI LSL+

Sbjct: 185 AVFLPTLLDAELFTKKQKTSALLSYLTLALVLITLSLM 222

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 865> which encodes the amino acid sequence <SEQ ID 866>. Analysis of this protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

INTEGRAL	Likelihood = -9.92	Transmembrane	214 - 230 (212 - 232)
INTEGRAL	Likelihood = -5.36	Transmembrane	135 - 151 (128 - 153)
INTEGRAL	Likelihood = -1.81	Transmembrane	101 - 117 (100 - 117)
INTEGRAL	Likelihood = -1.44	Transmembrane	183 - 199 (182 - 199)
INTEGRAL	Likelihood = -0.53	Transmembrane	166 - 182 (166 - 182)

30 ----- Final Results -----

bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

- 35 The protein has homology with the following sequences in the databases:

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus uberis]
 Identities = 72/128 (56%), Positives = 94/128 (73%)

40 Query: 106 FLLLYVLSGMGNNAFTFWLTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGKSYQT 165
 FLLLYVLSG+MGN T + TP VAAGASTSLFGLFAAIVV+ + G NQ LK +GKSYQT
 Sbjct: 95 FLLLYVLSGIMGNVLTFFTPHVVAAGASTSLFGLFAAIVVVGYFGHNQLLKSIGKSYQT 154

45 Query: 166 LIVVNLLMNLFMPNVSMAGHIGGVGGALLSIVFPTKMRVITVKKT KRM LALVSYGIILV 225
 LI++NL+MNLFMPNV + GH+GG +GGAL ++ PT + K ++ AL+SY ++
 Sbjct: 155 LIILNLVMNLFMPNVGIVGHGALGGALAAVFLPTLLDAELFTKKQKTSALLSYLTLAL 214

50 Query: 226 GVLVLGFL 233
 ++ L +
 Sbjct: 215 VLITLSLM 222

An alignment of the GAS and GBS proteins is shown below:

Identities = 63/132 (47%), Positives = 92/132 (68%)

55 Query: 94 GSLRFLILYILSGLMGNIIFTLFFTPHVVAAGASTSLFGLFAAIAIAGYFGKNPYLKQVGK 153
 G FL+LY+LSG+MGN FT + TP VAAGASTSLFGLFAAIVVLSFLGKNQALKDLGK
 Sbjct: 102 GLTPFLLLYVLSGMGNNAFTFWLTPETVAAGASTSLFGLFAAIVVLSFLGKNQALKDLGK 161

Query: 154 SYQVMILLNLFFNIFTPGVSLAGHVGGLVGGVVAIFLTKQNGSLLFKTWQSILALMIFI 213
 SYQ +I++NL N+F P VS+A GH+GG+VGG L++I + + K + +LAL+ +

60 Sbjct: 162 SYQTLIVVNLLMNLFMPNVSMAGHIGGVGGALLSIVFPTKMRVITVKKT KRM LALVSYG 221

Query: 214 IVSISLIGLSLV 225
 I+ + ++ L +
 Sbjct: 222 IIIVGVLVLGFL 233

A further corresponding DNA sequence was identified in *S.pyogenes* <SEQ ID 9083> which encodes the amino acid sequence <SEQ ID 9084>. Analysis of this protein sequence reveals the following:

Possible site: 52

5 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -7.70 Transmembrane 12 - 28 (7 - 30)

 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS sequences follows:

15 Score = 74.5 bits (180), Expect = 5e-16
 Identities = 37/96 (38%), Positives = 48/96 (49%)

 20 Query: 1 MTQLLKRYPXXXXXXXXXXXXAMQVYGHILATGAQAIYQVGGMFGLLVKAMPDQLWRL 60
 M + K YP MQ+ YG A +Q I+Q GG+ G +KA P LWRL
 Sbjct: 3 MKKFAKEYPTTVLLVSLTTLVFLMLQTYGSQAESSQVIFQFGGIQGDYLKAYPTNLWRL 62

 25 Query: 61 VTPXXXXXXXXXXXXVNGLTLFVVGQIVEDLWGSRLF 96
 ++P +NGL LYFVGQ+ E +WGS F
 Sbjct: 63 ISPIFVHIGWEHFLNGLALYFVGQMGESIWGSRLF 98

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 271

A DNA sequence (GBSx0296) was identified in *S.agalactiae* <SEQ ID 867> which encodes the amino acid sequence <SEQ ID 868>. Analysis of this protein sequence reveals the following:

Possible site: 22
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2055(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

40 >GP:BAA28715 GB:AB001562 hypothetical protein [Streptococcus mutans]
 Identities = 96/173 (55%), Positives = 129/173 (74%)

 45 Query: 1 MEKKLLRKEVLITLKSQPQAYKSEVDCKLLEAFIKTKAYQNCSVIATYLSFDYEYNTQLL 60
 M KK R +V+ LK Q +A K D +LLE I+ +AYQ + VIATYL+F +E++T LL
 Sbjct: 1 MMKKDYRTQVIEDLKKQDKAKKVLRDEQLLEELIQLEAYQKAHVIATYLAFTPFEFDTSLL 60

 50 Query: 61 IKQALCDGKRVLPKTKYPKGKMI FVDYQKDNLRRTPFGLLEPVNDRAVEKASIDLHVPG 120
 I+QA D K ++VPKTYP+ KMIFV Y + +L+ T FGL EP ++ A+EK++ IDLIHVPG
 Sbjct: 61 IEQAQRDNKSIVVPKTYPQRKMIFVVYDEADLQITKFGLKEPRSEEALEKSAIDLHVPG 120

 55 Query: 121 LIFNNKGFRIGYGAGYFDRYLSDFEGDTISTIYRCQRQDFVEEKHDVAVKEVL 173
 L FNN+G+RIG+GAGY+D+YL+DF+GDT+STIY Q+ F D+ VKEVL
 Sbjct: 121 LAFNNEGYRIGFGAGYYDQYLADFQGDTVSTIYSFQQFTFEPSFFDIPVKEVL 173

55 A related GBS nucleic acid sequence <SEQ ID 10925> which encodes amino acid sequence <SEQ ID 10926> was also identified.

-341-

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 272

- 5 A DNA sequence (GBSx0297) was identified in *S.agalactiae* <SEQ ID 869> which encodes the amino acid sequence <SEQ ID 870>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -1.44      Transmembrane 161 - 177 ( 161 - 177)
10   INTEGRAL      Likelihood = -0.22      Transmembrane 29 - 45 ( 28 - 45)

----- Final Results -----
      bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
15      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9305> which encodes amino acid sequence <SEQ ID 9306> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
20  >GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase
     [Streptococcus mutans]
     Identities = 344/401 (85%), Positives = 374/401 (92%)

25  Query: 1 MDLPENYDKEEFSRRIQKAAEKIKIKSDSEVLVVGIGGGSYLGAKAAIDFLNNHFANLQTAEE 60
     ++LP+NYDKEEF+RI+KAAEKKIKSDSEVLVVGIGGGSYLGAA+AAIDFLN+ F NL+ EE
     Sbjct: 49 LNLPQNYDKEEFAARIKKAAEKIKIKSDSEVLVVGIGGGSYLGARAAIDFLNSSFVNLENKEE 108

30  Query: 61 RKAQIILYAGNSISSTYLADLVEYVQDKEFSVNVISKGTTTEPAIAFRVFKEVVKKYG 120
     RKAQIILYAGNSISS YLADLV+YV DK+FSVNVISKGTTTEPAIAFRVF+LLVKKYG
     Sbjct: 109 RKAQIILYAGNSISSSYLAQDLYVADKDFSVNVISKGTTTEPAIAFRVFKEVLLVKKYG 168

35  Query: 121 QEEANKRIYATTDKVKGAVKVEADANNWETFVVPDNVGGRFSVLTAVGLLPIAASGADIT 180
     QEEAN+RIYATT+VKGAVKVEADAN WETFVVPD+VGGRF+VLTAVGLLPIAASGAD+
     Sbjct: 169 QEEANQRIYATTDKVKGAVKVEADANGWETFVVPDSVGGRFTVLTAVGLLPIAASGADLD 228

40  Query: 181 ALMEGANAAKDLSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWKQL 240
     LM GA AAR+D SS ++SEN AYQYAA+RN+LYRKGY+TE+LANYEPSLQYF EWWKQL
     Sbjct: 229 QLMAGAEEARQDYSSAELSENEAYQYAAIRNILYRKGYVTEVLANYEPSLQYFSEWWKQL 288

45  Query: 241 AGESEGKDQKGIYPTSANSFSTDHLHSLGQFIQEGYRNLFETVVRVEKPRKNVTIPELTEDL 300
     AGESEGKDQKGIYPTSANSFSTDHLHSLGQFIQEG RNLFETV+RVEK RKN+ +PE EDL
     Sbjct: 289 AGESEGKDQKGIYPTSANSFSTDHLHSLGQFIQEGNRNLFETVIRVEKARKNILVPEAAEDL 348

50  Query: 301 DGLGYLQGKDVFVNKKATDGVLLAHTDGGVPNMFVTLPTQDAYTLGYTTIYFFELAIGLS 360
     DGL YLQGKDVFVNKKATDGVLLAHTDGGVPN F+T+P QD +TLGY IYFFELAIGLS
     Sbjct: 349 DGLAYLQGKDVFVNKKATDGVLLAHTDGGVPNTFLTIPEQDEF TLGYTTIYFFELAIGLS 408

Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELSAELNARL 401
      GYLN VNPFDQPGVEAYK+NMFALLGKPGFEEL AELNARL
55  Sbjct: 409 GYLNGVNPFDQPGVEAYKKNMFALLGKPGFEELGAELNARL 449
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 871> which encodes the amino acid sequence <SEQ ID 872>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -1.44      Transmembrane 209 - 225 ( 209 - 225)
```

-342-

INTEGRAL Likelihood = -0.22 Transmembrane 77 - 93 (76 - 93)

----- Final Results -----

bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
 5 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAD33517 GB:AF132127 glucose-6-phosphate isomerase

[Streptococcus mutans]

Identities = 369/449 (82%), Positives = 408/449 (90%)

Query: 1 MSHTTFDYSKVLESFAGQHEIDFLQGVTEADKLLREGTGPBSDFLGWLIDLPENYDKDEF 60
 M+HI FDYSKVL F HE+D++Q QVT AD+ LR+GTGPG++ GWL+LP+NYDK+EF

15 Sbjct: 1 MTHIKFDYSKVLGKFLASHELDYIQMQVTAADEALRKGTGPGAEMTGWLNLQPQNYDKEEF 60

Query: 61 ARILTAAEKIKADSEVLVVIGIGGSYLGAKAAIDFLNHHFANLQTAKERKAPQILYAGNS 120
 ARI AAEKIK+DSEVLVVIGIGGSYLG+AIDFLN F NL+ +ERKAPQILYAGNS

20 Sbjct: 61 ARIKKAEEKIKSDSEVLVVIGIGGSYLGARAAGIDFLNSSFVNLENKEERKAPQILYAGNS 120

Query: 121 ISSTYLADLVEYVQDKEFSVNISKSGTTTEPAIAFRVFKEELVKKYGQEEANKRIYATT 180
 ISS YLADLV+YV DK+FSVNISKSGTTTEPAIAFRVF+LLVKKYGQEEAN+RIYATT

Sbjct: 121 ISSNYLADLVDYVADKDFSVNVISKSGTTTEPAIAFRVFKDLLVKKYGQEEANQRIYATT 180

25 Query: 181 DVKGAVKVEADANNWETFVVPDNVGGRFSVLTAVGLLPIAASGADITALMEGANAARKD 240
 D+VKGAVKVEADAN WETFVVPD+VGGRF+VLTAVGLLPIASGAD+ LM GA AAR+D

Sbjct: 181 DRVKGAVKVEADANGWETFVVPDSVGGRFTVLTAVGLLPIAASGADLDQLMAGAEAARQD 240

30 Query: 241 LSSDKISENIAYQYAAVRNVLYRKGYITEILANYEPSLQYFGEWWKQLAGESEGKDQKGI 300
 SS ++SEN AYQYAA+RN+LYRKGY+TE+LANYEPSLQYF EWWKQLAGESEGKDQKGI

Sbjct: 241 YSSAELSENAEYQYAAIRNLYRKGYTEVLANYEPSLQYFSEWWKQLAGESEGKDQKGI 300

Query: 301 YPTSANFSTDHLHSLGQFIQEGYRNLFETVIRVDNPRKNVIPELAEDLDGLGYLQGKDVD 360
 YPTSANFSTDHLHSLGQFIQEG RNLFETVIR+ RKN+++PE AEDLDGL YLQGKDVD

35 Sbjct: 301 YPTSANFSTDHLHSLGQFIQEGNRNLFETVIRVEKARKNIVPPEAAEDLDGLAYLQGKDVD 360

Query: 361 FVNKKATDGVLLAHTDGGVPNMFTLPAQDEFTLGYTIYFFELAIAVSGYMNADVNPFDQP 420
 FVNKKATDGVLLAHTDGGVPN F+T+P QDEFTLGY IYFFELAI +SGY+N VNPFQDQP

Sbjct: 361 FVNKKATDGVLLAHTDGGVPNTFLTIPEQDEFTLGYVIYFFELAIGLSGYLNGVNPFQDQP 420

40 Query: 421 GVEAYKRNMFALLGKPGFEALSAELNARL 449
 GVEAYK+NMFALLGKPGFE L AELNARL

Sbjct: 421 GVEAYKRNMFALLGKPGFEELGAELNARL 449

45 The protein has homology with the following sequences in the databases:

>GP:CAB90755 GB:AJ400707 hypothetical protein [Streptococcus
 uberis]

Identities = 58/91 (63%), Positives = 69/91 (75%)

50 Query: 6 KRYPTITIFLLGLTGLIFIAMQVYGHLATGAQAIYQVGGMFGLVKAMPDQLWRLVTPIF 65
 K P+T F L +T L+FI MQV YG A Q ++Q GGMFGL+VK+MP QLWRLVTPIF
 Sbjct: 5 KEKPVTFFFLLSVTILLFIVMQFYGSWAKSPQVVFQFGGMFGLVVKSMPSQLWRLVTPIF 64

55 Query: 66 IHIGFGHFFVNGLTLYFVGQIVEDLWGSRLF 96
 IHIG+ HF +N LTLYFVGQ+ E +WGSR F

Sbjct: 65 IHIGWEHFLINSLTLYFVGQLAESIWGSRFF 95

An alignment of the GAS and GBS proteins is shown below:

Identities = 380/401 (94%), Positives = 392/401 (96%)

60 Query: 1 MDLPENYDKEEFSRIQKAAEKIKSDSEVLVVIGIGGSYLGAKAAIDFLNNHFANLQTAAE 60
 +DL PENYDK+EF+RI AAEKIK+DSEVLVVIGIGGSYLGAKAAIDFLN+HFANLQTAA+E
 Sbjct: 49 LDLPENYDKDEFARILTAAEKIKADSEVLVVIGIGGSYLGAKAAIDFLNNHFANLQTAAE 108

-343-

Query: 61 RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKS GTTTEPAIAFRVF KELLVKKG 120
 RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKS GTTTEPAIAFRVF KELLVKKG
 Sbjct: 109 RKAPQILYAGNSISSTYLADLVEYVQDKEFSVNVISKS GTTTEPAIAFRVF KELLVKKG 168

5 Query: 121 QEEANKRIYATTDKVKGAVKVEADANNWETFVVVPDNVGGRFSVL TAVGLLPIAASGADIT 180
 QEEANKRIYATTDKVKGAVKVEADANNWETFVVVPDNVGGRFSVL TAVGLLPIAASGADIT
 Sbjct: 169 QEEANKRIYATTDKVKGAVKVEADANNWETFVVVPDNVGGRFSVL TAVGLLPIAASGADIT 228

10 Query: 181 ALMEGANAAKDLSSDKISENIA YQYA A VRNV LYRKGYITEI LANYEPLS LQYFGEWWKQL 240
 ALMEGANAAKDLSSDKISENIA YQYA A VRNV LYRKGYITEI LANYEPLS LQYFGEWWKQL
 Sbjct: 229 ALMEGANAAKDLSSDKISENIA YQYA A VRNV LYRKGYITEI LANYEPLS LQYFGEWWKQL 288

15 Query: 241 AGESEGKDQKG IYPT SANFSTD LHSL GQFI QEGYRNL FETV RVEKPRKNV TIPEL TEDL 300
 AGESEGKDQKG IYPT SANFSTD LHSL GQFI QEGYRNL FETV R+ RV+ PRKNV IPEL EDL
 Sbjct: 289 AGESEGKDQKG IYPT SANFSTD LHSL GQFI QEGYRNL FETV RDNPRKNV IIPELAEDL 348

20 Query: 301 DGLGYLQGKD VDFVNKKATDG VLLAHTDGGV PNMFVTLP TDAY T LGYTIYFFELAIGLS 360
 DGLGYLQGKD VDFVNKKATDG VLLAHTDGGV PNMFVTLP QD +TLGYTIYFFELAI +S
 Sbjct: 349 DGLGYLQGKD VDFVNKKATDG VLLAHTDGGV PNMFVTLP AQDEF TLGYTIYFFELAI AVS 408

Query: 361 GYLNSVNPFDQPGVEAYKRNMFALLGKPGFEELSAELNARL 401
 GY+N+VNPFDQPGVEAYKRNMFALLGKPGFE LSAELNARL
 Sbjct: 409 GYMNAVNPF DQPGVEAYKRNMFALLGKPGFEALSAELNARL 449

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 273

A DNA sequence (GBSx0298) was identified in *S.agalactiae* <SEQ ID 873> which encodes the amino acid sequence <SEQ ID 874>. Analysis of this protein sequence reveals the following:

30 Possible site: 38
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.66 Transmembrane 654 - 670 (653 - 671)
 INTEGRAL Likelihood = -1.65 Transmembrane 113 - 129 (113 - 129)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9463> which encodes amino acid sequence <SEQ ID 9464> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba histolytica]
 45 Identities = 536/864 (62%), Positives = 663/864 (76%), Gaps = 3/864 (0%)

Query: 20 ETTDVALAIDTLVQNGLKALDEMR--QLNQE QVDYIVAKASVAALDAH GELAL HAVEETG 77
 +T V I+ LV+ AL E + QE++DYIV KASVAALD H LA AVEETG
 Sbjct: 5 QTMTVDEHINQLVRKAQVALKEYLKPEYTQE KIDYIVKKASVAALDQHCALAAA VEEETG 64

50 Query: 78 RGV FEDKATKNLFACEHVNNMRHTKTVGVIEEDDV TGLT LIAEPVG VVCG ITPT TNPTS 137
 RG+FEDKATKNLFACEHV + MRH KTVG+I D + G+T IAEPVG VVCG+TP TNPTS
 Sbjct: 65 RGIFEDKATKNLFACEHVTHEMRHAKTVGIINVDPFLYGITEIAEPVG VVCG VTFVNPTS 124

55 Query: 138 TAIFKSLISLKLTRNPPI IFAFHPSAQESSAHAARI VRDAAIAAGAPENC VQWIEQPSIDAT 197
 TAIFKSLIS+KTRNPPI+F+FHPSA + S AA+JVRDAAIAAGAPENC+QWIE I+A+
 Sbjct: 125 TAIFKSLISIKTRNPPIVFSFHPSALKCSIMA KIVR DAAIAAGAPENCI QWIEFGGIEAS 184

60 Query: 198 NALMNHDGIATILATGGNAMVKAAYSCGKP ALGV GAGN VPAYEKSAN IRQA AH DIVMSK 257
 N LMNH G+ATILATGGNAMVKAAYS GKP ALGV GAGN VP Y+EK+ NI+QAA+D+VMSK

-344-

Sbjct: 185 NKLMNHPGVATILATGGNAMVKAAYSSGKPALGVGAGNVPTYIEKTCNIKQAANDVVMSK 244

Query: 258 SFDNGMVCASEQAVIIDKEIYKEFVEEFKSYHTYFVNKEKALLEEFCFGAKANSKNAG 317
SFDNGM+CASEQA IIDKEIY + VEE K+ YF+N+EKA LE+F FG A S +

5 Sbjct: 245 SFDNGMICASEQAAIIDKEIYDQVVEEMKTLGAYFINEEEKAKLEKFMFVGVNAYSADVNN 304

Query: 318 AKLNPNIVGKSAVWIAEQAGFTVPEGTNILAAECTEVSEKEPLTREKLSPVIAVLKAEST 377
A+LNP G S W AEQ G VPE NI+ A C EV EPLTREKLSPV+A+LKA+E+T

10 Sbjct: 305 ARLNPCKCPGMSPQWFAEQVGIKVPEDCNIICAVCKEVGPNEPLTREKLSPVLA ILKAENT 364

Query: 378 EDGVEKARQMVEFNGLGHSAAIHTKDADLAREFGTRIRAI RVIWNSPSTFGGIGDVYNAF 437
+DG++KA MVEFNG GHSAAIH+ D + ++ ++A R++ N+PS+ GGIG +YN

Sbjct: 365 QDGIDKAEMVEFNGRHSAAIHNSNDKAVV EKYLHNTPSSQGGIGSIYNYI 424

15 Query: 438 LPSLTLCGCSYGRNSVGDNVSAINLLNIKKVGRNNMWFQVKVPSKTYFERDSIQYLQKC 497
PS TLGCGSYG NSV NV+ NLLNIK++ RRNN+QWF+VP K +FE SI+YL +

Sbjct: 425 WPSFTLGCGSYGGNSVSANVTYHNLLNIKRLADRRNNLQWFRVPPKIFFEPHSIRYLAEL 484

20 Query: 498 RDVERVMIVTDHAMVELGFLDRITEQLDLRRNKVVYQIFAEVEPDPIITVMKGTDLMRT 557
+++ ++ IV+D M +LG++DR+++ L R N+V +IF +VEPDP I TV KG +M T

Sbjct: 485 KELSKIFIVSDRMMYKLGYVDRVMDVLKRRSNEVEIEIFIDVEPDPSIQTQKGLAVMNT 544

25 Query: 558 FKPDITIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGKKTFVA 617
F PD II A+GGGS MDAAK+MWL YE PE DF + QKF+D+RKRAFKFP +GKK + +

Sbjct: 545 FGPDNIIIAIGGGSAMDAKIMWLLYEHPEADFFAMKQKFIDLRKRAFKFPTMGKKARLIC 604

30 Query: 618 IPTTSGTGSEVTPFAVISDKANNRKYIADYSLTPPTVAIVDPALVMTVPGFIAADTGM DV 677
IPTTSGTGSEVTPFAVISD +KYP+ADYSLTP+VAIVDP M++P ADTG+DV

Sbjct: 605 IPTTSGTGSEVTPFAVISDHETGKKYPLADYSLTPSVAIVDPMFTMSLPKRAIADTGLDV 664

Query: 678 LHATEAYVSQMANDYTGLALQAIKIVFDYLERSVKDADFEAREKMHNASTMAGMAFAN 737
L HATEAYVS MAN+YTDGLA +A+K+VF+ L +S + D EAREKMHN+T+AGMAFA+

Sbjct: 665 LVHATEAYVSVMANEYTDGLAREAVKLFENLLKSY-NGDLEAREKMHNATIAGMAFAS 723

35 Query: 738 AFLGISHSHSMHKIGAQFHTVHGRTNAILLPYVIRYNGTRPAKTATWPKNYNYRADEKYQD 797
AFLG+ HSMHK+GA FH HGR A+LLP+VIRYNG +P K A WPKYN+Y+AD++Y +

Sbjct: 724 AFLGMDHSHMAHKVGAAFHLPHGRCVAVLLPHVIRYNGQKPRKLAMWPKYNFYKADQR YM 783

40 Query: 798 IAKLLGLPAATPPEAIVESYAKAVYDLGTRLGIKMNF RDQGIDEKEWKEKSRELAFLAYED 857
+A+++GL TP E VE++AKA +L F+ IDE W K B+A LA+ED

Sbjct: 784 LAQMVGGLKCNTPAEGVEAFAKACEELMKATEITITGFKKANIDEAAWMSKVPEMALLAFED 843

45 Query: 858 QCSPANPRLPMVDHMQEIIEDAYY 881
QCSPANPR+PMV M++I++ AYY

Sbjct: 844 QCSPANPRVPMVKDMEKILKAAYY 867

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 875> which encodes the amino acid sequence <SEQ ID 876>. Analysis of this protein sequence reveals the following:

Possible site: 55

50 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -3.66 Transmembrane 643 - 659 (642 - 660)
INTEGRAL Likelihood = -1.81 Transmembrane 102 - 118 (102 - 118)

55 ----- Final Results -----
bacterial membrane --- Certainty=0.2466 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

60 The protein has homology with the following sequences in the databases:

>GP:AAA81906 GB:U04863 alcohol dehydrogenase 2 [Entamoeba histolytica]
Identities = 535/870 (61%), Positives = 669/870 (76%), Gaps = 3/870 (0%)

65 Query: 6 NTVETTSVSVTIDALVQKGLAAL EEMRKLD--QE QV D YIVAKASVA ALDAH GELAKHAYE 63

-345-

Sbjct: 2 +T +T +V I+ LV+K AL+E K + QE++DYIV KASVAALD H LA A E
Sbjct: 2 STQQMTVDEHINQLVRKAQVALKEYLKPEYTQEKIDYIVKKASVAALDAHGCALAAA 61

5 Query: 64 ETGRGVFEDKATKHLFACEHVVNNMRHQKTVGIIIEEDVTGLTLIAEPVGVICGITPTTN 123
ETGRG+FEDKATK++FACEHV + MRH KTVGII D + G+T IAEPVGV+CG+TP TN
Sbjct: 62 ETGRGIFEDKATKNIFACEHVTHEMRHAKTVGIINVDPLYGITEIAEPGVVCGVTPVTN 121

10 Query: 124 PTSTAIFKSLISLKLTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCQWVETPSL 183
PTSTAIFKSLIS+KTRNPI+F+FHPSA + S AA+IVRDAIAAGAPENC+QW+E +
Sbjct: 122 PTSTAIFKSLISIKTRNPIVFSFHPSALKCSIMAAKIRVDAIAAGAPENCIQWIEFGGI 181

15 Query: 184 EATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAADIV 243
EA+N LMNH G+ATTILATGGNAMVKAAYS GKPALGVGAGNV Y+EK+ NI+QAA+D+V
Sbjct: 182 EASNKLMNHPGVATILATGGNAMVKAAYSSGKPALGVGAGNVPTYIEKTCNIKQAANDVV 241

20 Query: 244 MSKSFNDGMVCASEQAVIIDKEIYDDFVAEFSYHTYFVNKEKALLEFCFGAKANSKN 303
MSKSFNDGM+CASEQA IIDKEIYD V E K+ YF+N++EKA LE+F FG A S +
Sbjct: 242 MSKSFNDGMICASEQAAIIDKEIYDQVVEEMKTLGAYFINEEAKLEKFMFGVNAYSAD 301

25 Query: 304 CAGAKLNPNIVGKPATWIAEQAGFTVPEGTNILAAECKEVSENEPLTREKLSPVIAVLKS 363
A+LNP G W AEQ G VPE NI+ A CKEV NEPLTREKLSPV+A+LK+
Sbjct: 302 VNNARLNPKCPGMSPQWFAEQVGKVPEDCNIICAVCKEVGPNEPLTREKLSPVLAILKA 361

30 Query: 364 ESREDGVEKARQMVEFNGLGHSAAIHTADAELAKEFGTRIRAIRVIWNSPSTFGGIGDVY 423
E+ +DG++KA MVEFNG GHSAAIH+ D + +++ +A R++ N+PS+ GGIG +Y
Sbjct: 362 ENTQDGIDKAEAMVEFNGRHSAAIHSDKAVVEKYALTMACRILHNTPSQGGIGSIY 421

35 Query: 424 NAFLPSLTLCGGSYGRNAVDNVSAINLLNIKKVGRRRNNMQWFKVPSKTYFERDSIQYL 483
N PS TLGCGSYG N+V NV+ NLLNIK++ RRNN+QWF+VP K +FE SI+YL
Sbjct: 422 NYIWPSFTLGCSSYGGNSVSANVITYHNLLNIKRLADRRNNLQWFRVPPKIFFEPHSIRYL 481

40 Query: 484 QKCRDVERVMIVTDHAMVELGFLDRRIEQLDLRRNKVVYQIFAEVEPDDITTVMKGTTEL 543
+ +++ ++ IV+D M +LG++DR+++ L R N+V +IF +VEPDP I TV KG +
Sbjct: 482 AELKELSKIFIVSDRMMYKLGVYDRVMDDLVKRRSNEVEIEIFIDVEPDPSIQTQVKGLAV 541

45 Query: 544 MRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGKTK 603
M TF PD IIA+GGGS MDAAK+MWL YE PE DF + QKF+D+RKRAFKFP +GKK +
Sbjct: 542 MNTFGPDNIIAIGGGSAMDAAKIMWLLYEHPREADFFAMKQKFIDLRKRAFKFPTMGKKAR 601

50 Query: 604 FVAIPTTSGTGSEVTPFAVISDANKRKYPIADYSLTPTAIVDPALVLTVPGFIAADTG 663
+ IPTTSGTGSEVTPFAVISD +KYP+ADYSLTP+VAIVDP +++P ADTG
Sbjct: 602 LICIPPTTSGTGSEVTPFAVISDHEGKKYPLADYSLTPSVAIVDPMFTMSLPKRAIADTG 661

55 Query: 664 MDVLTHATEAYVSQMANDFTDGLALQAIKIVFDNLEKSVKTADFEAREKMHNASTMAGMA 723
+DVL HATEAYVS MAN++TDGLA +A+K+VF+NL KS D EAREKMHNAT+AGMA
Sbjct: 662 LDVLVHATEAYVSVMANEYTDGLAREAVKLFENLLKSY-NGDLEAREKMHNAAATIAGMA 720

60 Query: 724 FANAFLGISHSMAHKIGAQFHTVHGRNTAILLPYVIRYNGTRPAKTATWPKNYYRADEK 783
FA+AFLG+ HSMAHK+GA FH HGR A+LLP+VIRYNG +P K A WPKNY+Y+AD++
Sbjct: 721 FASAFLGMDHSMAHKVGAAFHLPHGRCVAVLLPHVIRYNGQKPRKLAMWPKNFYKADQR 780

65 Query: 784 YQDIAKLLGLPASTPEEAVESYAKAVYDLGCRVGIQMFKAQGIDENEWEKSRELAYLA 843
Y ++A+++GL +TP E VE++AKA +L FK IDE W E+A LA
Sbjct: 781 YMELAQMVGLKCNTPAEGVEAFAKACEELMATEITITGFKKANIDEAAWMSKVPEMALLA 840

70 Query: 844 YEDQCSPANPRLPMVDHMQEIIEDAYGYA 873
+EDQCSPANPR+PMV M++I++ AYY A
Sbjct: 841 FEDQCSPANPRVPMVKDMEKILKAAYYPIA 870

75 An alignment of the GAS and GBS proteins is shown below:

Identities = 827/880 (93%), Positives = 852/880 (95%)

Query: 12 MTEKTKAVETTDVALAIDTLVQNGLKALDEMROLNQEVDYIVAKASVAALDAHGEALAH 71
MTE VETT V++ ID LVQ GL AL+EMR+L+QEVDYIVAKASVAALDAHGEA H
Sbjct: 1 MTEGHNTVETTSVSVTIDALVQKGLAALEEMRKLDQEQVDYIVAKASVAALDAHGEALAH 60

80 Query: 72 AVEETGRGVFEDKATKNLFACEHVVNNMRHTKTVGIEEDDVTLTLIAEPGVVCGITP 131

-346-

A EETGRGVFEDKATK+LFACEHVNNMRH KTVG+IEEDDVTLGLTLIAEPVGV+CGITP
 Sbjct: 61 AYEETGRGVFEDKATKHLFACEHVNNMRHQKTVGIIIEEDDVTLGLTLIAEPVGVICGITP 120

5 Query: 132 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCVQWIEQ 191
 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCVQW+E
 Sbjct: 121 TTNPTSTAIFKSLISLKTRNPIIFAFHPSAQESSAHAARIVRDAIAAGAPENCVQWET 180

10 Query: 192 PSIDATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 251
 PS++ATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH
 Sbjct: 181 PSLEATNALMNHDGIATILATGGNAMVKAAYSCGKPALGVGAGNVPAYVEKSANIRQAAH 240

15 Query: 252 DIVMSKSFDNGMVCASEQAVIIDKEIYKEFVEEFKSYHTYFVNKKEKALLEEFCFGAKAN 311
 DIVMSKSFDNGMVCASEQAVIIDKEIY +FV EFKSYHTYFVNKKEKALLEEFCFGAKAN
 Sbjct: 241 DIVMSKSFDNGMVCASEQAVIIDKEIYDDFVAEFKSYHTYFVNKKEKALLEEFCFGAKAN 300

20 Query: 312 SKNCAGAKLNPNIVGKSAVWIAEQAGFTVPEGTNILAAECTEVSEKEPLTREKLSPVIAV 371
 SKNCAGAKLNPNIVGK A WIAEQAGFTVPEGTNILAAEC EVSE EPLTREKLSPVIAV
 Sbjct: 301 SKNCAGAKLNPNIVGKPATWIAEQAGFTVPEGTNILAAECKEVSENEPLTREKLSPVIAV 360

25 Query: 372 LKAESTEDGVEKARQMVFNGLGHSAAIHTKDADLAREFGTRIRAIRVIWNSPSTFGGIG 431
 LK+ES EDGVEKARQMVFNGLGHSAAIHT DA+LA+EFGTRIRAIRVIWNSPSTFGGIG
 Sbjct: 361 LKSESREDGVEKARQMVFNGLGHSAAIHTADELAKEFGTRIRAIRVIWNSPSTFGGIG 420

30 Query: 432 DVYNAFLPSLTLCGGSYGRNSVGDNSAINLLNIKKVGRRRNNMQWFVKVPSKYFERDSI 491
 DVYNAFLPSLTLCGGSYGRN+VGDNSAINLLNIKKVGRRRNNMQWFVKVPSKYFERDSI
 Sbjct: 421 DVYNAFLPSLTLCGGSYGRNAVGDNSAINLLNIKKVGRRRNNMQWFVKVPSKYFERDSI 480

35 Query: 492 QYLQKCRDVERVMIVTDHAMVELGFLDRIIIEQLDLRRNKVVYQIFAEVEPDPTTVMKG 551
 QYLQKCRDVERVMIVTDHAMVELGFLDRIIIEQLDLRRNKVVYQIFAEVEPDPTTVMKG
 Sbjct: 481 QYLQKCRDVERVMIVTDHAMVELGFLDRIIIEQLDLRRNKVVYQIFAEVEPDPTTVMKG 540

40 Query: 552 TDLMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK 611
 T+LMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK
 Sbjct: 541 TELMRTFKPDTIIALGGGSPMDAAKVMWLFYEQPEVDFHDLVQKFMDIRKRAFKFPELGK 600

45 Query: 612 KTKFVAIPPTSGTGEVTPFAVISDKANNRKYPIADYSLLPTVAIVDPALVMTVPGFIAA 671
 KTKFVAIPPTSGTGEVTPFAVISDKANNRKYPIADYSLLPTVAIVDPALV+TVPGFIAA
 Sbjct: 601 KTKFVAIPPTSGTGEVTPFAVISDKANNRKYPIADYSLLPTVAIVDPALVLTVPGFIAA 660

50 Query: 672 DTGMDVLTHATEAYVSQMANDYTDLALQAIIKIVFDYLERSVKDADFEAREKMHNASTMA 731
 DTGMDVLTHATEAYVSQMAND+TDGLALQAIIKIVFD LE+SVK ADFEAREKMHNASTMA
 Sbjct: 661 DTGMDVLTHATEAYVSQMANDFTDGLALQAIIKIVFDNLEKSVKTADFEAREKMHNASTMA 720

55 Query: 732 GMAFANAFLGISHSMAHKIGAQFHTVHGRNTAILLPYVIRYNGTRPAKTATWPKNYYRA 791
 GMAFANAFLGISHSMAHKIGAQFHTVHGRNTAILLPYVIRYNGTRPAKTATWPKNYYRA
 Sbjct: 721 GMAFANAFLGISHSMAHKIGAQFHTVHGRNTAILLPYVIRYNGTRPAKTATWPKNYYRA 780

Query: 792 DEKYQDIAKLLGLPAATPEEAVESYAKAVYDLGTRLGIKMFNRDQGIDEKEWKEKSRELA 851
 DEKYQDIAKLLGLPA+TPEEAVESYAKAVYDLG R+GI+MNF+ QGIDE EWKE SRELA
 Sbjct: 781 DEKYQDIAKLLGLPASTPEEAVESYAKAVYDLGCRVGIQMFKAQGIDENEWEKSRELA 840

Query: 852 FLAYEDQCSPANPRLPMVDHMQEIIIEDAYGYEERPGRRK 891
 +LAYEDQCSPANPRLPMVDHMQEIIIEDAYGY EERPGRRK
 Sbjct: 841 YLAYEDQCSPANPRLPMVDHMQEIIIEDAYGYAERPGRRK 880

A related GBS gene <SEQ ID 8533> and protein <SEQ ID 8534> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 10
 McG: Discrim Score: -4.68
 60 GvH: Signal Score (-7.5): -2.48
 Possible site: 21
 >>> Seems to have no N-terminal signal sequence
 ALOM program count: 1 value: -2.66 threshold: 0.0
 INTEGRAL Likelihood = -2.66 Transmembrane 100 - 116 (99 - 117)
 PERIPHERAL Likelihood = 3.61 173
 65 modified ALOM score: 1.03

*** Reasoning Step: 3

----- Final Results -----

5 bacterial membrane --- Certainty=0.2062 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8534 (GBS432) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell
10 extract is shown in Figure 173 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a His-fusion product.
SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 7; MW 41kDa).

GBS432-GST was purified as shown in Figure 223, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 274

A DNA sequence (GBSx0299) was identified in *S.agalactiae* <SEQ ID 877> which encodes the amino acid sequence <SEQ ID 878>. Analysis of this protein sequence reveals the following:

Possible site: 21

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3444 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 880.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

30 Example 275

A DNA sequence (GBSx0300) was identified in *S.agalactiae* <SEQ ID 881> which encodes the amino acid sequence <SEQ ID 882>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL	Likelihood = -8.39	Transmembrane	74 - 90	(69 - 94)
INTEGRAL	Likelihood = -5.31	Transmembrane	168 - 184	(163 - 186)
INTEGRAL	Likelihood = -4.83	Transmembrane	34 - 50	(29 - 52)
INTEGRAL	Likelihood = -0.75	Transmembrane	202 - 218	(202 - 219)

----- Final Results -----

bacterial membrane --- Certainty=0.4354 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA17305 GB:AL021926 hypothetical protein Rv0111 [Mycobacterium tuberculosis]

Identities = 70/218 (32%), Positives = 104/218 (47%), Gaps = 12/218 (5%)

45 Query: 9 VRITGLLLVLLYHFFKNSFPGGFVGVDIFFTFSGFLITALLIDEFSKKIDFVSFCRRR 68
 +R + LVL H GGF+GVD FF SGFLIT+LL+DE +T +ID F RR

-348-

Sbjct: 39 LRAIAVALVLASHGGIPGMGGGFIGVDAFFVLSGLITSLLLDELGRTGRIDLSGFWIRR 98

Query: 69 FYRIFPPPLVLMVLVTIIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGNYESQFI-P 127
R+ P LVLMVL L + S + A +T+N+ + +Y +Q P

5 Sbjct: 99 ARRLLPALVLMVLTVSAARALFPDQALTGLRSDAIAAFLWTANWRFVAQNTDYFTQGAPP 158

Query: 128 HLFVHTWSLSIEVHFYVLWGL---TVWLLSKRSKDQKQLRGTLFLISMGIFGVSFITMF 183
HTWSL +E +YV+W L LL+ R++ ++ R T+ + F ++ L

10 Sbjct: 159 SPLQHTWSLGVEEQYYVVWPPLLIGATLLAARAR-RRCRRATVGGVRFAAFLIASLGT 217

Query: 184 VRAFFVDNFST-----IYFSTLSHIFPFILGAMVATI 215

A F++ IYF T + +G+ A +

Sbjct: 218 ASATAAVAFSAATRDRIYFGTDTRAQALLIGSAAAAL 255

15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 879> which encodes the amino acid sequence <SEQ ID 880>. Analysis of this protein sequence reveals the following:

Possible site: 46

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood = -10.83 Transmembrane 325 - 341 (313 - 346)

20 INTEGRAL Likelihood = -9.29 Transmembrane 237 - 253 (234 - 258)

INTEGRAL Likelihood = -7.91 Transmembrane 166 - 182 (162 - 188)

INTEGRAL Likelihood = -6.10 Transmembrane 72 - 88 (68 - 92)

INTEGRAL Likelihood = -4.09 Transmembrane 264 - 280 (260 - 281)

INTEGRAL Likelihood = -2.87 Transmembrane 371 - 387 (370 - 390)

25 INTEGRAL Likelihood = -2.66 Transmembrane 34 - 50 (32 - 50)

INTEGRAL Likelihood = -1.91 Transmembrane 3 - 19 (3 - 19)

INTEGRAL Likelihood = -0.85 Transmembrane 136 - 152 (136 - 154)

----- Final Results -----

30 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

35 Identities = 167/226 (73%), Positives = 195/226 (85%)

Query: 1 MRIKWFSLVRITGLLLVLLYHFKNSFPGGFVGVDIFFTSGLITALLIDEFSKKID 60
MRIKWFS VR+TGLLLVLLYHFKFN FPAGF+GVDIFFTSGLITALLIDE++K + ID

40 Sbjct: 1 MRIKWFSFVRTGLLLVLLYHFKKNVFPGGFIGVVDIFFTSGLITALLIDEYTKKESID 60

Query: 61 FVSFCRRRFYRIFPPPLVLMVLVTIIPFVFLVKSDFRASIGSQIMTALGFTSNFYEILTGGN 120
+ F +RRFYRII PPLVLM+L+TIPF FL+K DF A+IGSQI LGFT+N YEILTG +

Sbjct: 61 IIIGFLKRRRFYRIVPPPLVLMILITIPFTFLIKKDFIANIGSQITAVLGFTTNIYEILTGS 120

45 Query: 121 YESQF1PHLFVHTWSLSIEVHFYVLWGLTVLLSKRSKDQKQLRGTLFLISMGIFGVSL 180
YESQF1PHLFVHTWSL+IEVHFY+ WG+ VWLL++R + QKQLRG LFLIS+GIF +SFL

Sbjct: 121 YESQF1PHLFVHTWSLAIIEVHFYLFWGVFVWLARRKETQKQLRGFLISLGIFAISFL 180

Query: 181 TMFVRAFFVDNFSTIYFSTLSHIFPFILGAMVATISGIRETTGRFK 226

+MF+R+F NFS IYFS+LSH PPPFLGAM ATI+GI E T RF+

Sbjct: 181 SMFIRSFMTSNFSLIYFSSLHSFPFFLGAMFATITGINETTVRFQ 226

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 276

A DNA sequence (GBSx0302) was identified in *S.agalactiae* <SEQ ID 883> which encodes the amino acid sequence <SEQ ID 884>. Analysis of this protein sequence reveals the following:

Possible site: 38

>>> Seems to have a cleavable N-term signal seq.

-349-

----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

!GB:AE004818 hypothetical protein [Pseudomonas aerug...
 !GB:AE004818 hypothetical protein [Pseudomonas aerug...

10

>GP:AAG07403 GB:AE004818 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 33/80 (41%), Positives = 50/80 (62%)

15

Query: 45 KYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGSYTGDFKKQPDGQ 104
 +Y G +V+ + G+G+L Y+NG +Y G F +G+ G GT+ G Y+G F G DGQ
 Sbjct: 39 RYRGELVDGRLEGQGRLDYDNGAWYAGRFEHGLLHGHGTWQGADGSRYSGGFAAGLFDGQ 98

20

Query: 105 GRLNAKNKKVYKGTFKQGIY 124
 GRL + VY+G F+QG++
 Sbjct: 99 GRLAMADGSVYQGGFRQGLF 118
 Identities = 31/91 (34%), Positives = 46/91 (50%), Gaps = 2/91 (2%)

25

Query: 34 QGVFSYDGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGSYT 93
 QG YD G Y G + + G G +G Y G F G+F+G+G G Y
 Sbjct: 52 QGRLDYDNGAW-YAGRFEHGLLHGHGTWQGADGSRYSGGFAAGLFDGQGRALAMADGSVYQ 110

30

Query: 94 GDFKKQPDGQGRNLNAKNKKVYKGTFKQGIY 124
 G F++G DG+G L + + Y+G F++G+Y
 Sbjct: 111 GGFRQQLFDGEGSLEQQGTR-YRGGFRKGLY 140
 Identities = 31/91 (34%), Positives = 42/91 (46%), Gaps = 1/91 (1%)

35

Query: 32 SSQGVFSYDGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGS 91
 S QG G +Y GS + G+G + G+ Y G F +G GKG + G
 Sbjct: 141 SGQGTLGSDGS-RYQGSFRQGRLEGEGSFSDSQGNQYAGTFRDGQLNGKGRWSPGDGR 199

40

Query: 92 YTGDFKKQPDGQGRNLNAKNKKVYKGTFKQG 122
 Y G FK Q GQGR + + V+ G F +G
 Sbjct: 200 YVCQFKDNQFHGQGRYESASGDVWIGRFSEG 230
 Identities = 31/91 (34%), Positives = 45/91 (49%), Gaps = 4/91 (4%)

45

Query: 34 QGVFSYDGK---IKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHG 89
 QG+F +G +Y G +G+G L +G Y+G F G EG+G+F G
 Sbjct: 115 QGLFDGEGSLEQQGTRYRGFRKGLYSQGQTLGSDGSRYQGSFRQGRLEGEGSFDSQG 174

50

Query: 90 WSYTGDFKKQPDGQGRNLNAKNKKVYKGTFK 120
 Y G F+ GQ +G+GR + + Y G FK
 Sbjct: 175 NQYAGTFRDGQLNGKGRWSPGDGRYVGQFK 205
 Identities = 28/87 (32%), Positives = 45/87 (51%), Gaps = 1/87 (1%)

55

Query: 34 QGVFSYDGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGSYT 93
 +G FS G +Y G+ + + GKG+ + +GD Y G F + F G+G + S G +
 Sbjct: 166 EGSSFSDSQGN-QYAGTFRDGQLNGKGRWSPGDGRYVGQFKDNQFHGQGRYESASGDVWI 224

60

Query: 94 GDFKKQPDGQGRNLNAKNKKVYKGTFK 120
 G F +G +G G L + Y+G F+
 Sbjct: 225 GRFSEGALNGPGEELLGADGSRYRGFFQ 251
 Identities = 28/89 (31%), Positives = 43/89 (47%), Gaps = 2/89 (2%)

65

Query: 34 QGVFSYDGKIKYVGSIVNHHMTGKGKLTYENGDYYKGDFVNGVFEGKGTFVSVHGSYT 93
 QG + G + Y G G+G L + G Y+G F G++ G+GT G Y
 Sbjct: 98 QGRALAMADGSV-YQGGFRQGLFDGEGSLE-QQGTRYRGFRKGLYSQGQTLGSDGSRYQ 155

Query: 94 GDFKKQPDGQGRNLNAKNKKVYKGTFKQG 122
 G F++G+ +G+G + Y GTF+ G
 Sbjct: 156 GSFRQGRLEGEGSFSDSQGNQYAGTFRDG 184

-350-

Identities = 25/80 (31%), Positives = 37/80 (46%)

Query: 45 KYVGSIVNHHMTGKGKLTENGDYYKGDFVNNGVFEKGKTFVSVHGSYTGDFKKQGPDGQ 104
+YVG ++ G+G+ +GD + G F G G G + G Y G F+ + GQ

5 Sbjct: 199 RYVGQFKDNQFHGQGRYESASGDVWIGRFSEGALNGPGELLGADGSRYRGGFQFWRFHQ 258

Query: 105 GRLNAKNKKVYKGTFKQGIY 124
G L + Y+G F G Y

Sbjct: 259 GLLEQLDGTRYEGGFAAGAY 278

10

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 885> which encodes the amino acid sequence <SEQ ID 886>. Analysis of this protein sequence reveals the following:

Possible site: 35

15 >>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood =-13.16 Transmembrane 20 - 36 (12 - 41)

----- Final Results -----

bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA16606 GB:D90899 hypothetical protein [Synechocystis sp.]
25 Identities = 37/89 (41%), Positives = 49/89 (54%), Gaps = 6/89 (6%)

Query: 48 KGRMHYT-----GYVINHKMNGEGLVLPNGDIYEGTFKDGLFEGKGTFTAKTGWLNG 101
KG YT G V+ ++NG GK Y NGD YEGT K+G +G+G F G Y G
Sbjct: 141 KGTIFIYTNGDRCSGTVVQGELNGSGKCEYNNGDQYEGTLKNGQPDGEGIFRFAAGGEYEG 200

30

Query: 102 EFHKGQANGKGVLKAKNNKVKYKGIFKQGI 130
EF G+ +G+G N ++G FKQG+
Sbjct: 201 EFQSGEFGQQGTRIFANGNRFQGQFKQGL 229

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 68/126 (53%), Positives = 93/126 (72%)

Query: 1 MKNFKITRTHLEILSIIIVFGLSVFTLTTSSQGVFSYDGGKIKYVGSIVNHHMTGKGK 60
+K + ITR LEI+S+I+I+V +SVF++ S++ +YD G++ Y G ++NH M G+GK
40 Sbjct: 8 VKKWSITRAKLEIVSVIVILVCAISVFSVRISNKTSLYDKGRMHTGYVINVHKMNGEK 67

Query: 61 LTYENGDYYKGDFVNNGVFEKGKTFVSVHGSYTGDFKKQGQGRILNAKNNKVKYKGTFK 120
L Y NGD Y+G F +G+FEKGKTF + GW Y G+F KGQ +G+G L AKN KKYKG FK
Sbjct: 68 LVYPNGDIYEGTFKDGLFEGKGTFTAKTGWLNGEFHKGQANGKGVLKAKNNKVKYKGIFK 127

45

Query: 121 QGIYQK 126
QGI+QK
Sbjct: 128 QGIFQK 133

50 SEQ ID 884 (GBS139) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 3; MW 13kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 22 (lane 2; MW 38.2kDa), in Figure 24 (lane 7; MW 38kDa) and in Figure 33 (lane 7; MW 38.2kDa).

The GBS139-GST fusion product was purified (Figure 200, lane 2) and used to immunise mice. The 55 resulting antiserum was used for FACS (Figure 287), which confirmed that the protein is immunoaccessible on GBS bacteria.

-351-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 277

A DNA sequence (GBSx0303) was identified in *S.agalactiae* <SEQ ID 887> which encodes the amino acid

5 sequence <SEQ ID 888>. This protein is predicted to be holliday junction dna helicase ruvb (ruvB). Analysis of this protein sequence reveals the following:

Possible site: 59

>>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4386 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB75331 GB:Y15896 RuvB protein [Bacillus subtilis]
 Identities = 196/322 (60%), Positives = 254/322 (78%)

20 Query: 3 RFLDS DAMG D EEL VERT LRP QV LRE YIG QD KV K DQL K I FIE AAK LR D E S LD H V L L FG PPG 62
 R + S+A E ++E++L RPQ L + YIG Q KV K+ L++F I+A AK+R E+LD H V L L +G PPG
 Sbjct: 4 RL VS SEAD NH ES VIE QSL RPQ NLA QY I G QH KV KEN L RV FIDA AAK MRQ ET LD H V L L Y G PPG 63

Query: 63 LGKTT MAFVIAN ELGV NL KQT SGPA IE KSG D L V AIL ND L E PG D V L FIDE I HR MP MA VE EV 122
 LGKTT+A ++A NE+G V L+ T SG PA IE+ G D L A IL L E PG D V L FIDE I HR+ ++EEV
 25 Sbjct: 64 LGKTT LASI V AN EM G VEL RTT SG PA IE R PG D L A IL T A L E PG D V L FIDE I HRL H RS IE EV 123

Query: 123 LYS AMED FYI D I MIGA GETS RSV HLD L P PFT LIG AT TRAG MLS NPL RAR F G IT GH ME YY E 182
 LY AMED F +DI+IG G ++RSV LD L P PFT L +GAT TR G +L+ PL R RFG+ +EYY
 Sbjct: 124 LY PAM EDF C L DIV I G K GPS A RSV RL D L P PFT L V GAT T R V G L L T A P L R D R F G V M S R LE YY T 183

30 Query: 183 END L TE II I ERT AD I FEM KITY EA ASEL A RRS RGT P RIA N RLL K R V RD Y A Q I MG D GL I D DN 242
 + +L + I+ RT AD+ FE++I + A E+ ARR S RGT P R+ A N RLL+ R V RD+ A Q++ GD I ++
 Sbjct: 184 Q EEL A D I V T R T A D V F E V E I D K P S A L E I A R R S RGT P R V A N R L L R R V D F A Q V L G D S R I T E D 243

35 Query: 243 IT DK AL T M L D V D H E G L D Y D Q K I L R T M I E M Y N G G P V G L G T L S V N I A E E R D T V E D M Y E P Y L 302
 I+ AL L V D G L D++D K+L M I E +N G G P V G L T+S I E E T+ED+Y E P Y L
 Sbjct: 244 I S Q N A L E R L Q V D R L G L D H I D H K L L M G M I E K F N G G P V G L D T I S A T I G E E S H T I E D V Y E P Y L 303

40 Query: 303 I Q K G F I M R T R T G R V A T V K A Y E H 324
 +Q G F I R T G R+ T Y H
 Sbjct: 304 L Q I G F I Q R T P R G R I V T P A V Y H H 325

A related GBS nucleic acid sequence <SEQ ID 10943> which encodes amino acid sequence <SEQ ID 10944> was also identified.

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 889> which encodes the amino acid sequence <SEQ ID 890>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0686 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 An alignment of the GAS and GBS proteins is shown below:

Identities = 282/327 (86%), Positives = 306/327 (93%)

-352-

```

Query: 1 MTRFLDSAMGDEELVERTLRPQYLREYIGQDKVKDQLKIFIEAAKLRDESLDHVLLFGP 60
        M R LD++ MG+EE +RTLRPQYL EYIGQDKVK+Q IFIEAAK RDESLDHVLLFGP
Sbjct: 25 MARILDNNVMGNEEFSRTLRLPQYLHEYIGQDKVKKEQFAIFIEAAKRRDESLDHVLLFGP 84

5 Query: 61 PGLGKTTMAFVIANELGVNLKQTSGPAIEKSGDLVAILNDLEPGDVLFIDEIHRMPMAVE 120
        PGLGKTTMAFVIANELGVNLKQTSGPA+EK+GDLVAILN+LEPGD+LFIGDEIHRMPM+VE
Sbjct: 85 PGLGKTTMAFVIANELGVNLKQTSGPAVEKAGDLVAILNELEPGDILFIGDEIHRMPMSVE 144

10 Query: 121 EVLYSAMEDFYIDIMIGAGETSRSVHLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY 180
        EVLYSAMEDFYIDIMIGAG+TSRS+HLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY
Sbjct: 145 EVLYSAMEDFYIDIMIGAGDTSRSTHLDLPPFTLIGATTRAGMLSNPLRARFGITGHMEY 204

15 Query: 181 YEENDLTEIIERTADIFEMKITYEASELARRSRGTPRIANLLKRVRDYAQIMGDLID 240
        Y+E DLTEII+ERTA IFE+KI +EAA +LA RSRGTPRIANLLKRVRDYAQI+GDG+I
Sbjct: 205 YQEKKDLTEIERTATIFELIKIDHEAARKLACRSRGTPRIANLLKRVRDYAQIIGDGII 264

20 Query: 241 DNITDKALTMLDVDHEGLDYVDQKILRTMIEMYNGGPVGLGTL SVNIAEERDTVEDMYEP 300
        ITD+ALTMLDVD EGLDY+DQKILRTMIEMY GGPVGLGTL SVNIAEER+TVE+MYEP
Sbjct: 265 AQITDRALTMULDVDRGELDYIDQKILRTMIEMYQGGPVGLGTL SVNIAEERNTVEEMYEP 324

Query: 301 YLIQKGFIMRTRTGRVATVKAYEHLGY 327
        YLIQKGF+MRTRTGRVAT KAY HLGY
Sbjct: 325 YLIQKGFLMRTGRTGRVATQKAYRHILGY 351

```

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 278

A DNA sequence (GBSx0304) was identified in *S.agalactiae* <SEQ ID 891> which encodes the amino acid sequence <SEQ ID 892>. Analysis of this protein sequence reveals the following:

```

30 Possible site: 43
    >>> Seems to have no N-terminal signal sequence
        INTEGRAL Likelihood = -2.87 Transmembrane 157 - 173 ( 157 - 174)
        INTEGRAL Likelihood = -1.49 Transmembrane 205 - 221 ( 205 - 222)

35 ----- Final Results -----
        bacterial membrane --- Certainty=0.2147(Affirmative) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

40 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 893> which encodes the amino acid sequence <SEQ ID 894>. Analysis of this protein sequence reveals the following:

```

Possible site: 56
>>> Seems to have no N-terminal signal sequence
45 ----- Final Results -----
        bacterial cytoplasm --- Certainty=0.3097(Affirmative) < succ>
        bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

50 An alignment of the GAS and GBS proteins is shown below:

Identities = 130/303 (42%), Positives = 202/303 (65%)

```

55 Query: 1 MLKHFGSKVRNLRVTRNITREDFCGDETELSVRQLARIESGQSIPNLTKAHYIAKQLNVK 60
        ML+HFG KV+ LR+ + I+RED CGDE+ELSVRQLARIE GQSIP+L+K +IAK LNV
Sbjct: 1 MLEHFGGKVVLRLERKRISREDLCGDESELVRQLARIELGQSIPSLSKVIFIAKALNVS 60

Query: 61 LDILTGGESELPLPKRYKELKYLILRIPTYADAERLKLRECQFDHIFEELYDNLPEDECLA 120
        + LT G LELPKRYKELKYLILR PTY D +L++RE QFD IFE++YD LPE+E +

```

Sbjct: 61 VGYLTDGADLELPKRYKELKYLILRTPTYMDDGKLQVREEQFDEIFEDYYDKLPEEKII 120

Query: 121 IDSLQAKFEVYQTGDINFGEVLCCECFDKVKYKEKYTLNDLIIIDFLTCAVVSKFNNRA 180
ID LQA + + + NFG++L E F++K K ++ NDLI++L+L + + +

5 Sbjct: 121 IDCLQATLDTLLSENTNFGIDLLQEYFNQIKTKVFRQNDLILLELYLAYLDIEGMGQY 180

Query: 181 FTKEVFQTIKTLISQNHKLTAEDLFWFNHVLLNCVFVGLCLNSEECLAEMLEVSQTMV 240
K + ++ L Q + ++LF N + + + + L N + L + +E+S++ M

10 Sbjct: 181 SDKIFYDSLLDNLSEQFEQFELDELFIVNKIIIDISSLKNNRLDNLEKAIEMSQKIMA 240

Query: 241 STHDFHKMPLYFMYWQKYFITIDNDIKSAENAYQQSIMFSKMIDDKHLIKKLELEWQEDI 300
D+++MP+ + +WKYF+ DI AE ++ ++ +F++M D++L KL EW++D+

Sbjct: 241 KIQDWNRMPILKLIIEWKYFLIKQKDIKAEQSFMKACLFQAQMKTADQYLENKLIQEWEKD 300

15 Query: 301 TGH 303

+

Sbjct: 301 KSY 303

SEQ ID 892 (GBS319) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 4; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 7; MW 62kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 279

25 A DNA sequence (GBSx0305) was identified in *S.agalactiae* <SEQ ID 895> which encodes the amino acid sequence <SEQ ID 896>. This protein is predicted to be adenylosuccinate lyase (purB). Analysis of this protein sequence reveals the following:

Possible site: 35

>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3358 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04344 GB:AP001509 adenylosuccinate lyase [Bacillus halodurans]
Identities = 326/430 (75%), Positives = 366/430 (84%)

40 Query: 1 MIERYSRPEMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60
MIERY+RPEM AIWTEEN+Y+AWLEVEI+A EAWAELGEIPKEDV KIRE A FD++RIL
Sbjct: 1 MIERYTRPEMGAIWTEENRYQAWLEVEIIVACEAWAELGEIPKEDVKKIREHASFDVERIL 60

45 Query: 61 EIEQDTRHDVVAFTRAVSETLGEEERKWHYGLTSTDVVDTAYGYLYKQANDIIRRDLNF 120
EIEQ+TRHDVVAFTRAVSETLGEEERKWHYGLTSTDVVDTA YL KQAN+II DL F
Sbjct: 61 EIEQETRHDVVAFTRAVSETLGEEERKWHYGLTSTDVVDTALSYLLKQANEIIEADLVRF 120

50 Query: 121 TNIVADAKEHKFTIMMGRTHGVHAEPPTFGLKLATWYSEMRNTERFEHAAAGVEAGKI 180
+I+ +KA EHKA+T+MMGRTHGVHAEPPTFGLKLA WY EMKRN+ERF AA GV GK+
Sbjct: 121 LDILKEKALEHKYTVMMSGRTGVHAEPPTFGLKLALWYEEMKRNLERFRLAEGVRVGKL 180

55 Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240
SGAVG +ANI PFVEQYVC+KLG+ ISTQ L RD HAEY A LA IATSIE+ A EI
Sbjct: 181 SGAVGTYANIDPFVEQYVCEKLGLERAPISTQTLQRDRHAEYMATLALIATSIEKFAVEI 240

Query: 241 RGLQKSEQREVEEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVAYENVALWHE 300
RGLQKSE REVEE+FAKGQKGSSAMPHKRNPIGSENMTG+ARV+RGHM+ AYENV LWHE
Sbjct: 241 RGLQKSETREVEEYFAKGQKGSSAMPHKRNPIGSENMTGIARVVRGHMLAAAYENVPLWHE 300

-354-

Query: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360
 RDISHSSAERII PD TI I+YMLNRFGNIVKNLTVFPENM RNM T+GLIYSQRV+L
 Sbjct: 301 RDISHSSAERIILPDATIAINYMLNRFGNIVKNLTVFPENMKRNMTRTYGLIYSQRVLLS 360

5 Query: 361 LIEKGMTREEAYDLVQPKTAYSDNQVDFKPLLEEDTKVTSCLTQEEIDELNPPIYYTKR 420
 LI+KGM REEAYDLVQPK +W+ V F+ L+E++ ++TS L+ EEI+ F+ ++ K
 Sbjct: 361 LIDKGMVREEAYDLVQPKAMEAWEKGVQFRELVEQERITSVLSPEEIEACFDYNHHLKH 420

10 Query: 421 VDDIFERLGL 430
 VD IFERLGL
 Sbjct: 421 VDTIFERLGL 430

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 897> which encodes the amino acid sequence <SEQ ID 898>. Analysis of this protein sequence reveals the following:

15 Possible site: 35
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.3358 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 422/430 (98%), Positives = 428/430 (99%)

25 Query: 1 MIERYSRPEMAAIWTEENKYRAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60
 M+ERYSRPEMAAIWTEENKY AWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL
 Sbjct: 1 MLERYSRPEMAAIWTEENKYHAWLEVEILADEAWAELGEIPKEDVAKIREKADFDIDRIL 60

30 Query: 61 EIEQDTRHDVVAFTRAVSETLGEERKWKHYGLTSTDVVDTAYGYLYKQANDIIRRDIENF 120
 EIEQDTRHDVVAFTRAVSETLGEERKWKHYGLTSTDVVDTAYGYLYKQANDIIRRDIENF
 Sbjct: 61 EIEQDTRHDVVAFTRAVSETLGEERKWKHYGLTSTDVVDTAYGYLYKQANDIIRRDIENF 120

35 Query: 121 TNIVADKAKEHKFTIMMGRTHGVHAEPPTFGLKLATWYSEMKRNIERFEHAAAGVEAGKI 180
 TNIVADKA+EHK TIMMGRTHGVHAEPPTFGLKLATWYSEMKRNIERFEHAAAGVEAGKI
 Sbjct: 121 TNIVADKAREHKMTIMMGRTHGVHAEPPTFGLKLATWYSEMKRNIERFEHAAAGVEAGKI 180

40 Query: 181 SGAVGNFANIPPFVEQYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240
 SGAVGNFANIPPFVE+YVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI
 Sbjct: 181 SGAVGNFANIPPFVEEYVCDKLGIRPQEISTQVLPRDLHAEYFAVLASIATSIERMATEI 240

45 Query: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVALWHE 300
 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENV+LWHE
 Sbjct: 241 RGLQKSEQREVEEFFAKGQKGSSAMPHKRNPIGSENMTGLARVIRGHMVTAYENVSLWHE 300

50 Query: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360
 RDISHSSAERII ITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK
 Sbjct: 301 RDISHSSAERIITPDTTILIDYMLNRFGNIVKNLTVFPENMMRNMESTFGLIYSQRVMLK 360

55 Query: 361 LIEKGMTREEAYDLVQPKTAYSDNQVDFKPLLEEDTKVTSCLTQEEIDELNPPIYYTKR 420
 LIEKGMTREEAYDLVQPKTAYSDNQVDFKPLLEEDTKVTSCLTQEEIDELNPPIYYTKR
 Sbjct: 361 LIEKGMTREEAYDLVQPKTAYSDNQVDFKPLLEEDTKVTSCLTQEEIDELNPPIYYTKR 420

Query: 421 VDDIFERLGL 430
 VDDIF+RLG+
 Sbjct: 421 VDDIFKRLGI 430

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-355-

Example 280

A DNA sequence (GBSx0306) was identified in *S.agalactiae* <SEQ ID 899> which encodes the amino acid sequence <SEQ ID 900>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 45
      >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -16.24 Transmembrane 145 - 161 ( 119 - 167)
      INTEGRAL Likelihood = -9.98 Transmembrane 125 - 141 ( 119 - 144)
      INTEGRAL Likelihood = -9.29 Transmembrane 28 - 44 ( 23 - 51)
      INTEGRAL Likelihood = -7.01 Transmembrane 196 - 212 ( 193 - 220)
10     INTEGRAL Likelihood = -6.21 Transmembrane 96 - 112 ( 88 - 116)
      INTEGRAL Likelihood = -5.79 Transmembrane 249 - 265 ( 246 - 266)
      INTEGRAL Likelihood = -2.87 Transmembrane 222 - 238 ( 222 - 238)
      INTEGRAL Likelihood = -2.28 Transmembrane 279 - 295 ( 278 - 295)

15 ----- Final Results -----
      bacterial membrane --- Certainty=0.7496(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

20 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB13498 GB:AB028634 RNA polymerase [Flammulina velutipes]
  Identities = 83/336 (24%), Positives = 150/336 (43%), Gaps = 40/336 (11%)
```

```

25 Query: 152 ILLLIAFVSIGKNR-VYNFVQNLNYFEEVIWNYFEENPVKIKEKSЛИK-----FLLTIS 205
          IL L   SI NR +  ++ N    ++ N+F+ + +K   K L+I       F++ +S
          Sbjct: 133 ILLFLYLIYSILINRFILKWLDNSGIIYKININWFKNHMIKHINKMLVINIKFFNFIKLS 192

30 Query: 206 FVFVIDFAMVRL-----LNFNKFSTTILACSAILLAWLYQN-----KSVTEPFL 249
          + +I  +++ L    +NF+I+  I   I   ++           S+   F
          Sbjct: 193 IITIIGISIMELFGIFGINFDIRIIINYLKINTSGKJHLTTINMDQYSVLENSIHTIFY 252

35 Query: 250 LKKLVIVYFIFIATLIGNLKN-ELSILETPLLFISIFFTMDRRIIALSKEMRDLI--ISKS 306
          + L+I+ IF   L  N+KN + +I   +L+I IF   I   ++DL+  ++K
          Sbjct: 253 INLLIIFLIFISLILYRNVKNIDTNIKRWIILYIIFLINIIFIFNHIYIKDLMNDLNKY 312

40 Query: 307 ILFYDYHENIKPSILLSEIKEIKYLENDIGE---LELVRQMVIRLRLLEEEFLILSDI 363
          IL Y D   I  S+ L    ++K L+ ++I +   V+ + I+ ++E   L +  I
          Sbjct: 313 IILDYMDLHIIIVNSLFLFNKFDVK-LKRINIYKSYSTVTVKDLEIKSKIEERSNELDIKLI 371

45 Query: 364 YMKNNG-YEKYIQFVQGNVYFINL--LDKIPNYTNLKLILESIFD---HNNQKIFIPKL 416
          K G YE YI  ++ N+  ++ E L   P Y N   +E + +   +   F+ K+
          Sbjct: 372 IAKYGSYENYINSIE-NINIVDEEFILKNYPEYINDSKFIEFLMELEPLFRDHTEFVKKI 430

Query: 417 YEEYIYIYLISLGEVEKAKEIL---KEVSDYLTEESL 449
45      YE        L +  K+IL   KE+ DY+ + +L
      Sbjct: 431 YENLNINSTNEKLEFLLANKDILSENKEIFDYVLQLNL 466
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for 50 vaccines or diagnostics.

Example 281

A DNA sequence (GBSx0308) was identified in *S.agalactiae* <SEQ ID 901> which encodes the amino acid sequence <SEQ ID 902>. Analysis of this protein sequence reveals the following:

```

55 Possible site: 37
      >>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3307(Affirmative) < succ>
```

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 282

- A DNA sequence (GBSx0309) was identified in *S.agalactiae* <SEQ ID 903> which encodes the amino acid
 10 sequence <SEQ ID 904>. This protein is predicted to be purK (purK). Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

- 15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 A related GBS nucleic acid sequence <SEQ ID 9461> which encodes amino acid sequence <SEQ ID 9462> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA04376 GB:AJ000883 purK [Lactococcus lactis]
 Identities = 208/347 (59%), Positives = 258/347 (73%), Gaps = 3/347 (0%)

- 25 Query: 14 NSFKTIGIIGGGQLGQMMIAAAIYMGHKVITLDPASDCPASRVS-EVIVAPYDDVEALGT 72
 N+ +TIGIIGGGQLGQMMIAAA YMGHKVITLDP +C A++VS E+IVAPYDDVE L
 Sbjct: 4 NTKQTIGIIGGGQLGQMMIAAAQYMGHKVITLDPNPNCASAKVSDDELIVAPYDDVENLLR 63
- 30 Query: 73 LAARCDVLTYEFENVADGLDAVVSAGQLPQGTDLLRISONRIFEKFDFLANKAGVTVAPY 132
 LA CDV+TYEFENV A L + ++PQG LL I+QNR FEK+FL N+A V VAP+
 Sbjct: 64 LAYACDVTYEFENVSAKALHEIEGCVRIPQGIRLLEITQNRFEKEFLTNEAKVNVPW 123
- 35 Query: 133 KVVTSSSLDEGLDLTKTYVLKTATGGYDGHGQKVIRSAEDLPEAQQLANSAQCVLEEFVN 192
 ++V S+ L +T+ VLKT TGGYDGHGQ V+ + E L A+ L ++CVLE+F++
 Sbjct: 124 QLVDSAELKPET-VTRKQVLKTTGGYDGHGQVVLNTDEKLSAAKSLTTELSECVLEDFIS 182
- 40 Query: 193 FDLEISVIVSGNGQDVTVFPVQENIHRNNILSKTIVPARISDQLADKAKEMAVQIAKKLQ 252
 F+ EISVI+SGNG + VFP+ EN HR NIL +TI PARIS ++ + A ++A IA+KL+
 Sbjct: 183 FEREISVIISGNGHEYVFPLAENEHRENILHQTISSPARISAEITENAYKIATSTIAEKLE 242
- 45 Query: 253 LSGTLCVEMFATAD-DITVNEIAPRPHNSGHYSIEACDFSQFDTHILGVVLGAPLPIKLH 311
 LSG LCVEMF TAD I VNE+APRPHNSGH++IEACDF+QFD HI G+LG LP KL
 Sbjct: 243 LSGVLCVEMFLTADGQIYVNEIAPRPHNSGHFTIEACDFNQFDLHIKGILGEDLPEPKLL 302
- Query: 312 APAVMFNVLGQHVQQAIDHVQAQNPSAHLHMYGKLEAKHNRKMGHVT 358
 PA+M NVLGQHV+ ++ H H YGK +AKHNRKMGHVT+
 Sbjct: 303 KPAIMLNVLGQHVVEAVKKLNHEADWHQHDYKGADAKHNRKMGHVTI 349

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 905> which encodes the amino acid sequence <SEQ ID 906>. Analysis of this protein sequence reveals the following:

Possible site: 34
 >>> Seems to have no N-terminal signal sequence

- 55 ----- Final Results -----

-357-

```
bacterial cytoplasm --- Certainty=0.0334 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 344/369 (93%), Positives = 353/369 (95%)

```
Query: 1 MRNKEKSQRSQAMNSFKTIGIIGGGQLGQMMAIAAAIYMGHKVITLDPASDCPASRVSEVI 60
       MRNKEKSQRSQ +NSFKTIGIIGGGQLGQMMAIAAAIYMGHKVITLDPASD PASRVSEVI
10 Sbjct: 1 MRNKEKSQRSQVVNSFKTIGIIGGGQLGQMMAIAAAIYMGHKVITLDPASD PASRVSEVI 60

Query: 61 VAPYDDVEALGTLAARCDVLTYEFENVADGLDAVVSAGQLPQGTDLRISQNRIFEKDF 120
       VAPYDDVEALG LAARCDVLTYEFENVADGLDAVVS QLPQGTDLRISQNRI EKDF
15 Sbjct: 61 VAPYDDVEALGQLAARCDVLTYEFENVADGLDAVVSACQLPQGTDLRISQNRIVEKDF 120

Query: 121 LANKAGVTVAPYKVVTSLLDEGLDLTKTYVLKTATGGYDGHGQKVIRSAEDLPEAQQLA 180
       LANKAGVTVAPYKVVTSLL GLDLTKTYVLKT TGGYDGHGQK+IRSAEDLPEAQQLA
Sbjct: 121 LANKAGVTVAPYKVVTSLLDGGLDLTKTYVLKTETGGYDGHGQKIIIRSAEDLPEAQQLA 180

20 Query: 181 NSAQCVCVLEEFVNFDLEISVIVSGNGQDVTVFPVQENIHRNNILSKTIVPARISDQLADKA 240
       NSAQCVCVLEEFVNFDLEISVIVSGNG+DVTVFPVQENIHRNNILSKTIVPARISDQLADKA
Sbjct: 181 NSAQCVCVLEEFVNFDLEISVIVSGNGKDVTVFPVQENIHRNNILSKTIVPARISDQLADKA 240

25 Query: 241 KEMAVQIAKKLQLSGTLCVEMFATADDIIVNEIAPRPHNSGHYSIEACDFSQFDTHILGV 300
       K+ AVQIAKKLQLSGTLCVEMF TADDIIVNEIAPRPHNSG YSIEACDFSQFDTHILGV
Sbjct: 241 KKTAVQIAKKLQLSGTLCVEMFTTADDIIVNEIAPRPHNSGRYSIEACDFSQFDTHILGV 300

Query: 301 LGAPLPIPKLHAPAVMFNVLGQHVQQAIDHVAQNPSAHLHMYKLEAKHNRKMGHVTVFS 360
       LGAPLP I+LHAPAVM NVLGQHVQQA D+VA+NPSAHLHMYKLEAKHNRKMGHVTVF+
30 Sbjct: 301 LGAPLPQIQLHAPAVMLNVLGQHVQQATDYVAKNPSAHLHMYKLEAKHNRKMGHVTVFA 360

Query: 361 DVPDEVEEF 369
       DEV+EF
Sbjct: 361 KDADEVKEF 369
35
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 283

A DNA sequence (GBSx0310) was identified in *S.agalactiae* <SEQ ID 907> which encodes the amino acid sequence <SEQ ID 908>. This protein is predicted to be phosphoribosylaminoimidazole carboxylase catalytic subunit (purE). Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3572 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

50 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12462 GB:Z99107 phosphoribosylaminoimidazole carboxylase I
[Bacillus subtilis]
Identities = 106/162 (65%), Positives = 128/162 (78%)
```

```
55 Query: 33 MQPIIISIIMGSKSDWTTMQKTAEVLDNFIAYEKKVVAHRTPDLMFKHAEEARGRGIKI 92
       MQP++ IIMGS SDW TM+ ++LD + YEKKVVAHRTPD MF++AE AR RGIK+
Sbjct: 1 MQPLVGIIMGSTSDWETMKHACDILDELNVPYEKVVSAHRTPDFMFYEAETARERGIKV 60

Query: 93 IIAGAGGA AHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVATMAIGEAG 152
       IIAGAGGA AHLPGM AAKTTLPVIGVPV+S+AL+G+DSL SIVQMPGGVPVAT +IG+AG
60
```

-358-

```

Sbjct: 61 IIAGAGGAAHLPGMTAAKTTLPVIGVPVQSKALNGMDSLLSIVQMPGGPVATTSIGKAG 120
Query: 153 ATNAALTALRILSTEDQNLADALAHFHEEQGKIAEESSNELI 194
      A NA L A +ILS D++LA L E + ESS++L+
5 Sbjct: 121 AVNAGLAAQILSAFDEDLARKLDERRENTKQTVCLESDQLV 162

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 909> which encodes the amino acid sequence <SEQ ID 910>. Analysis of this protein sequence reveals the following:

```

Possible site: 57
10    >>> Seems to have no N-terminal signal sequence
        INTEGRAL Likelihood = -3.08 Transmembrane 36 - 52 ( 34 - 52)

----- Final Results -----
15      bacterial membrane --- Certainty=0.2232(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

>GP:CAA04375 GB:AJ000883 purE [Lactococcus lactis]
20    Identities = 105/158 (66%), Positives = 131/158 (82%)

Query: 46 ISIIMGSKSDWATMQKTAEVLDNFIAYEKKVVAHRTPDLMFKHAEEARGRGKIIAG 105
      ++IIMG SDWATM++TA++LD+FG+AYEKKVVAHRTPLM++ +AR RG K+IIAG
Sbjct: 4 VAIIMGCSSDWATMKETAKILDFFGLAYEKVVVAHRTPALMAEFSSQARERGYKVIIAG 63
25
Query: 106 AGGA AHL PGM VAA KTT LPV IGV PV KSR AL S GLD S L Y S IV QM PGG VP VAT MA IGE A GAT NA 165
      AGGA AHL PGM V+A+T +PVIGVP+KSR AL S GLD S L Y S IV QM PGP VAT MA IGE A G A NA
Sbjct: 64 AGGA AHL PGM VSA QTL VPV IGV PIK SR AL S GLD S L Y S IV QM PAG VP VAT MA IGE A G A K NA 123
30
Query: 166 ALT AL RIL SIE DQ NL AD AL AH FHE EQ G KIA E ESS GEL I 203
      AL AL++L+ ++NL L + ++ EES+ L+
Sbjct: 124 ALF AL QLL AN T NEN LIQ KLL VY RAA QEM VEE SNK ALL 161

```

An alignment of the GAS and GBS proteins is shown below:

```

35    Identities = 162/169 (95%), Positives = 164/169 (96%), Gaps = 1/169 (0%)

Query: 27 PLYLNIMQ-PIISIIMGSKSDWTTMQKTAEVLDNFIAYEKKVVAHRTPDLMFKHAEEA 85
      PL + IM+ PIISIIMGSKSDW TMQKTAEVLDNFIAYEKKVVAHRTPDLMFKHAEEA
Sbjct: 35 PLCILIMKTPIISIIMGSKSDWATMQKTAEVLDNFIAYEKKVVAHRTPDLMFKHAEEA 94
40
Query: 86 RGRGIKIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVAT 145
      RGRGIKIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVAT
Sbjct: 95 RGRGIKIIIAGAGGAAHLPGMVAAKTTLPVIGVPVKSRALSGLDSLYSIVQMPGGVPVAT 154
45
Query: 146 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSNELI 194
      MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESS ELI
Sbjct: 155 MAIGEAGATNAALTALRILSIEDQNLADALAHFHEEQGKIAEESSGELI 203

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for 50 vaccines or diagnostics.

Example 284

A DNA sequence (GBSx0311) was identified in *S.agalactiae* <SEQ ID 911> which encodes the amino acid sequence <SEQ ID 912>. This protein is predicted to be phosphoribosylglycinamide synthetase (purD). Analysis of this protein sequence reveals the following:

```

55    Possible site: 16
        >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```

-359-

bacterial cytoplasm --- Certainty=0.1966 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 298/419 (70%), Gaps = 7/419 (1%)

10 Query: 1 MKLLVVGSGGREHAIAKKLLASKDQVFVAPGNDGMLTDGLDLVNIGISEHSRLIDFVK 60
 MK+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ F +
 Sbjct: 1 MKILVIGSGGREHALAKKFMEQPVEEVFVAPGNSGMEDGQIQIVHISELSNDKLVKFAQ 60

15 Query: 61 ENEIAWTLIGPDDALAAGIVDGFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 120
 I T +GP+ AL G+VD F A L FGP K AAELE SKDFAK IM KY VPTA
 Sbjct: 61 NQNIGLTFVGPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

20 Query: 121 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 180
 Y TF E A AY++E+G P+V+KADGLA GKGV VA +E A A ++ F S
 Sbjct: 121 YATFDSLEPALAYLDEKGVPVLVIKADGLAAGKGVTVAFDIETAKSALADI----FSGSQ 175

25 Query: 181 ARVVIEEFLDGEESFLFAFANGDKFYIMPTAQDHKRAYDGDKGLNTGGMGAYAPVPHLPQ 240
 +VVIEEFLDGEESFLF+F + K Y MP AQDHKRA+D DKG NTGGMGAY+PV H+ +
 Sbjct: 176 GKVVIEEFLDGEESFLFSFIHDGKIYPMPIAQDHKRAFDENDKGPNNTGGMGAYSPVHLISK 235

30 Query: 241 SVVDTAVETIVKPVLEGPMIAEGRPYLGVLYAGLILITADGPVIEFNSRGDPETQIILPR 300
 VV+ A+E +VKP + GMI EG+ + GVLYAGLILIT DG K IEFN+RFGDPETQ++LPR
 Sbjct: 236 EVVNEALEKVVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295

35 Query: 301 LTSDFAQNIDDIMMGIIEPYITWQKDGVTLGVVVASEGYPLDYEKGVPLPEKTDGDIITYYY 360
 L SD AQ I DI+ G EP + W + GVTGVVVA+EGYP + G+ LPE +G + YY
 Sbjct: 296 LKSDLQAQAIIDILIAGNEPTLEWLESGVTLGVVVAEGYPSQAKLGLLPEIPEG-LNVYY 354

40 Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAI 419
 AG EN++ L+S+GGRVY++ T + VK+ Q +Y +L+ + G FYR+DIGS+AI
 Sbjct: 355 AGVSKNENNQ-LISSLGGRVYLVSETGEDVKSTQKLLYKEQLKLENDGFFYRHDIGSRAI 412

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 913> which encodes the amino acid sequence <SEQ ID 914>. Analysis of this protein sequence reveals the following:

Possible site: 35
 40 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.80 Transmembrane 5 - 21 (5 - 21)

45 ----- Final Results -----
 bacterial membrane --- Certainty=0.1319 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

50 Query: 50 LKLLVVGSGGREHAIAKKLLASKDQVFVAPGNDGMLTDGLDLVNIVVSEHSRLIAFAK 109
 +K+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ FA+
 Sbjct: 1 MKILVIGSGGREHALAKKFMEQPVEEVFVAPGNSGMEDGQIQIVHISELSNDKLVKFAQ 60

55 Query: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169
 I F+GP+ AL G+VD F A L FGP K AAELE SKDFAK IM KY VPTA
 Sbjct: 61 NQNIGLTFVGPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

60 Query: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229
 Y TF E A AY++E+G P+V+KADGLA GKGV VA +E A A ++ F S
 Sbjct: 121 YATFDSLEPALAYLDEKGVPVLVIKADGLAAGKGVTVAFDIETAKSALADI----FSGSQ 175

Query: 230 ARVVIEEFLDGEESFLFAFANGDKFYIMPTAQDHKRAFDGDKGPNTGGMGAYAPVPHLPQ 289

-360-

+VVIEEFLDGEESLRF+F + K Y MP AQDHKRAFD DKGPNTGGMGAY+PV H+ +
Sbjct: 176 GKVVIEEFLDGEESLFSFIHDGKIYPMPIAQDHKRAFDKGPNTGGMGAYSPVLHISK 235

5 Query: 290 SVVDTAVEMIVRPVLEGMAEGRPYLGVLVGLILITADGPVIEFNSRFGDPETQIILPR 349
VV+ A+E +V+P + GM+ EG+ + GVLY GLILT DG K IEFN+RGDPETQ++LPR
Sbjct: 236 EVVNEALEKVVKPTVAGMIEEGKSFTGVLYAGLILITEDGVKTIEFNARFGDPETQVVLPR 295

10 Query: 350 LTSDFAQNIDIMMGIIEPYITWQDGVTLGVVVASEGYPDYEKGVPLPEKTDGDIITYY 409
L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP + G+ LPE +G + YY
Sbjct: 296 LKSDLQAQTIDILAGNEPTLEWLESGVTLGVVVAEEGYPSSQAKLGLLPEIPEG-LNVYY 354

15 Query: 410 AGVKFSENELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDDTTGLFYRNDIGSKAI 468
AGV +EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI
Sbjct: 355 AGVSKNENNQ-LISSLGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFFYRHDIGSRAI 412

An alignment of the GAS and GBS proteins is shown below:

Identities = 399/421 (94%), Positives = 408/421 (96%)

20 Query: 1 MKLLVVSGGREHAIAKKLASKDQDVQFVAPGNDGMTLDGLDLVNIGISEHSRLIDFVK 60
+KLLVVSGGREHAIAKKLASKDQDVQFVAPGNDGMTLDGLDLVN +SEHSRLI F K
Sbjct: 50 LKLLVVSGGREHAIAKKLASKDQDVQFVAPGNDGMTLDGLDLVNIVVSEHSRLIAFAK 109

25 Query: 61 ENEIAWTLIGPDDALAAGIVDGFSAGLRAFPGPTKAAAELEWSKDFAKEIMVKYNVPTAA 120
ENEI+W IGPDDALAAGIVD FNSAGLRAFPGPTKAAAELEWSKDFAKEIMVKYNVPTAA
Sbjct: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFPGPTKAAAELEWSKDFAKEIMVKYNVPTAA 169

30 Query: 121 YGTFSDFEKAKAYIEEQAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSC 180
YGTFSDFEKAKAYIEEQAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSC
Sbjct: 170 YGTFSDFEKAKAYIEEQAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSC 229

35 Query: 181 ARVVIEEFLDGEESLFAFANGDKFYIMPTAQDHKRAYDGDKGLENNTGGMGAYAPVPHLPQ 240
ARVVIEEFLDGEESLFAFANGDKFYIMPTAQDHKRA+DGDKG NTGGMGAYAPVPHLPQ
Sbjct: 230 ARVVIEEFLDGEESLFAFANGDKFYIMPTAQDHKRAFDGDKGPNTGGMGAYAPVPHLPQ 289

40 Query: 241 SVVDTAVETIVKPVLEGMAEGRPYLGVLVGLILITADGPVIEFNSRFGDPETQIILPR 300
SVVDTAVE IV+PVLEGMAEGRPYLGVLV GLILTADGPVIEFNSRFGDPETQIILPR
Sbjct: 290 SVVDTAVEMIVRPVLEGMAEGRPYLGVLVGLILITADGPVIEFNSRFGDPETQIILPR 349

45 Query: 301 LTSDFAQNIDIMMGIIEPYITWQDGVTLGVVVASEGYPDYEKGVPLPEKTDGDIITYY 360
LTSDFAQNIDIMMGIIEPYITWQDGVTLGVVVASEGYP DYBKGVPLPEKTDGDIITYY
Sbjct: 350 LTSDFAQNIDIMMGIIEPYITWQDGVTLGVVVASEGYPDYEKGVPLPEKTDGDIITYY 409

Query: 361 AGAKFAENSKALLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDDTTGLFYRNDIGSKAIKE 421
AG KF+ENS+ LLNGGRVYMLVTTEDSVKAGQDKIYTQLAQDDTTGLFYRNDIGSKAI+E
Sbjct: 410 AGVKFSENELLSSNGGRVYMLVTTEDSVKAGQDKIYTQLAQDDTTGLFYRNDIGSKAIKE 470

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 285

50 A DNA sequence (GBSx0312) was identified in *S.agalactiae* <SEQ ID 915> which encodes the amino acid sequence <SEQ ID 916>. Analysis of this protein sequence reveals the following:

Possible site: 36
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -1.28 Transmembrane 235 - 251 (235 - 251)

55 ----- Final Results -----
bacterial membrane --- Certainty=0.1510(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

60 The protein has homology with the following sequences in the GENPEPT database:

-361-

>GP:AAA23257 GB:M81878 unknown [Clostridium perfringens]
 Identities = 66/258 (25%), Positives = 119/258 (45%), Gaps = 9/258 (3%)

5 Query: 1 MTIYDQIESALDLMTDLEREIACYFMGQPISKDALASTIVTKQLHISQAALTRFAKKCGF 60
 M I +Q+E+ T E+ + Y + + +I+ K+ + +A +TRF KK GF
 Sbjct: 1 MGILEQLENPKFKATKSEKTLIEYIKSDLDNTIYKSISIIAKESGVGEATITRFTKKLG 60

10 Query: 61 KGYREFVFEYLKS-HETISQQLYGLQNDNTKKVFMNYQEMISKSADI-----IDEQL 112
 G+++F K + + L + V +M+ S +I ID + +
 Sbjct: 61 NGFQDFKVTLAKEISNKKNTSIINLHVHRDESVTETANKMLKSSINILEQTVKQIDL 120

15 Query: 113 LEVSHMIEQADRVYFYKGSSSLVAKEFKIRLMRLGVVICEALDDTDSFSWTSIVNDRCL 172
 + +I A RVYF G G S + A + + MR+G + D+ + +SI ND +
 Sbjct: 121 CKCRDLIMNAKRVYFIGIGYSGIAATDINYKFMRIGFTTVPTDSHTMVIMSSITND 180

20 Query: 173 VIAFSLSGNTNSVIGALKIASCHGAKTVLFTK-QPHTDYAFDKIIQVASARHLDYGNRI 231
 ++A S SG T VI +K A +G K + T+ + + D + SA + I
 Sbjct: 181 IVAISNSGTTKEVIKTVQAKENGTKIITLTEDSDNPLRKLSDYELTYTSAETIFETGS 240

25 Query: 232 SPQIPMLIMVDIIYAQFL 249
 S +IP + ++D++Y + +
 Sbjct: 241 SSKIPQIFLLDLLYTEVI 258

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 917> which encodes the amino acid sequence <SEQ ID 918>. Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.88 Transmembrane 243 - 259 (242 - 261)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.2954 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35 A related sequence was also identified <SEQ ID 9093> which encodes the amino acid sequence <SEQ ID 9094>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 56
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.88 Transmembrane 239 - 255 (238 - 257)

40 ----- Final Results -----
 bacterial membrane --- Certainty= 0.295 (Affirmative) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

45 An alignment of the GAS and GBS proteins is shown below:

Identities = 138/263 (52%), Positives = 189/263 (71%), Gaps = 2/263 (0%)

50 Query: 6 QIESALDLMTDLEREIACYFMGQPISKDALASTIVTKQLHISQAALTRFAKKCGFKGYRE 65
 +IE++L+ MT LE+ IA +F+ ++ L ++ + K+LHISQAALTRFAKKCGF GYR
 Sbjct: 14 KIEASLEHMITSLEKGIAHFFITDLTPQELTASEIVKRLHISQAALTRFAKKCGFTGYRA 73

55 Query: 66 FVFEYLKSHE TISQQLYGLQNDNTKKVFMNYQEMISKSADI IDEEQLLEVSHMIEQADRV 125
 F F+YL S + + + TK+V M+Y +I+K+ + + +EE+LL ++ +I+ ++RV
 Sbjct: 74 FAFDYLHSLQESQETFQSIIHLTKRVLMDYDALINKTYELVNEEKLLNLAKLIDSSERV 133

60 Query: 126 YFYKGSSSLVAKEFKIRLMRLGVVICEALDDTDSFSWTSIVNDRCLVIAFSLSGNTNSV 185
 YF+GKSS LVA+E K+R MRLG+IC+A DTD F+W NS+VN+ CLV FSLSG TNSV
 Sbjct: 134 YFFGKGSSGLVAREMKLRFMRGLICDAYSDTDGFTWANSLVNENCLVFGFSLSGKTNSV 193

Query: 186 IGALKIASCHGAKTVLFTKQPHT-IDYAFDKIIQVASARHLDYGNRISPQIPMLIMVDII 244
 I AL AS GAKTVL T T D + D II V+S L YGNR+SPQ P+LIM+DII
 Sbjct: 194 ITALHQASORGAKTVLTTDNQTEFDDSLD-IIIPVSSTHQLHYGNRVSPQFPLLIMMDII 252

Query: 245 YAQFLDINKIEKERIFRETIIQR 267

YA L I+K KE+IF+ TII +

Sbjct: 253 YAYVLAIDKPHKEKIFKNTIIDK 275

5

SEQ ID 916 (GBS320) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 40 (lane 5; MW 33kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 7; MW 58kDa) and in Figure 160 (lane 7 & 8; MW 58kDa).

10 GBS320-GST was purified as shown in Figure 224, lane 3-4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 286

15 A DNA sequence (GBSx0313) was identified in *S.agalactiae* <SEQ ID 919> which encodes the amino acid sequence <SEQ ID 920>. This protein is predicted to be xylan esterase 1 (cephalosporin-C). Analysis of this protein sequence reveals the following:

Possible site: 54
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.4981(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAB68821 GB:AF001926 xylan esterase 1 [Thermoanaerobacterium sp.
 'JW/SL YS485']
 Identities = 133/299 (44%), Positives = 188/299 (62%), Gaps = 1/299 (0%)

30 Query: 5 MSLDDMREYLGQDQIPEDFDDFWKKQTMKYQG-NIEYRLDKDFNITFAQAYDLHFKGSN 63
 M L +REY G + PEFD++W + + + + L + F ++FA+ YDL+F G
 Sbjct: 6 MPLQLREYTGTNPSCPEDFDEYWNRALDEMRSVDPKIELKESSFQVSFAECYDLYFTGVR 65

35 Query: 64 NSIVYAKCLFPKTNKPYPPVVFYFHGYQNQSPDWSDQQLNYVAAGYGVVSMMDVRGQAGQSQD 123
 + ++AK + PKT +P + FHGY + S DW+D+LNYVAAG+ VV+MDVRGQ GQSQD
 Sbjct: 66 GARIHAKYIKPKTEGKHPALIRFHGYSSNSGWDNDKLNYVAAGFTVVAMDVRGQGGQSQD 125

40 Query: 124 KGHFDGITVKQIVRGMISGPNHLFYKDIYLDVFQLIDIIATLESVDSNQLYSYGVWSQGG 183
 G G T+ G I+RG+ +++ ++ I+LD QL I+ + VD +++ G SQGG
 Sbjct: 126 VGGVTGNTLNIGHIRGLDDDADNMLFRHIFLDTAQLAGIVMMPEVDEDRVGVGMGPSQGG 185

45 Query: 184 ALALIAAALNPKIVKTVAVYPFLSDFRRVLDLGGVSEPYDELFRYFKYSDPFHKTEENNVL 243
 L+L AAL P++ K V+ YPFLSD++RV DL Y E+ YF+ DP H+ EN V
 Sbjct: 186 GLSLACAALEPRVRKVSEYPFLSDYKRVWDLIDLAKNAYQEITDYFRLFDPRHERENEVF 245

Query: 244 KTLAYIDVKNAFAHRISCPVVLITALKDDICPPSTQFAIFNRLTSTKKHILLPDYGHDPM 302
 L YIDVKN A RI V++ L D +CPPST FA +N + S K + PDYGH+PM
 Sbjct: 246 TKLGYIDVKNLAKRIKGDVLMCVGLMDQVCPNSTVFAAYNNIQSKKDIKVYPDYGHEPM 304

50 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

-363-

Example 287

A DNA sequence (GBSx0314) was identified in *S.agalactiae* <SEQ ID 921> which encodes the amino acid sequence <SEQ ID 922>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 35
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -5.73      Transmembrane 128 - 144 ( 126 - 145)

----- Final Results -----
10      bacterial membrane --- Certainty=0.3293 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:AAA23256 GB:M81878 unknown [Clostridium perfringens]
    Identities = 78/160 (48%), Positives = 110/160 (68%)

    Query: 131 CLTIGTGIGGCLIIDKTVFHGSNSACEVGYMHLSGDFQDLASTTALIADVAKAHGDEI 190
            CLTIGTGIGG LIID V HGFSNSA E+GYM ++ + QD+AS +AL+ +VA G E
    Sbjct: 18 CLTIGTGIGGALIIDGKVHLHGFSNSAGEIGYMMVNGENIQDIASASALVKNVALRKGVEP 77

20    Query: 191 SRWDGRRIFQEAKKGNEKCIAISIDRMINYLQGGIANMVYVVNPEKVVLLGGGIMAQKDYLQ 250
            S DGR + + G+ C +++++ + L GI+N+VY++NPE VVLGGGIMA+++ +
    Sbjct: 78 SSIDGRRYVLDNYENGDLICKEEVEKLADNLALGISNIVVYLINPEVVVLLGGGIMAREEVFR 137

25    Query: 251 DKLSESLKRNLTSLAEKTAIVFAQHENQAGMLGAYYHFK 290
            + SL++ L+ S+ T I FA+ +N AGM GAYY+FK
    Sbjct: 138 PLIENSIRKYLIESVYNNTKIAFAKLKNTAGMKGAYYNFK 177
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 923> which encodes the amino acid sequence <SEQ ID 924>. Analysis of this protein sequence reveals the following:

```

Possible site: 22
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -4.30      Transmembrane 128 - 144 ( 127 - 145)
    INTEGRAL      Likelihood = -0.11      Transmembrane 227 - 243 ( 227 - 243)

----- Final Results -----
        bacterial membrane --- Certainty=0.2720 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
40        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```

45 >GP:BAB04516 GB:AP001509 glucose kinase [Bacillus halodurans]
    Identities = 97/291 (33%), Positives = 155/291 (52%), Gaps = 14/291 (4%)

    Query: 5 LAIDIGGTAIKYGLISETGDLLEKEEMATEAYKGGPSILEVKGLVKTYQDQMDLAGVAI 64
            + ID+GGT IK L+S+ G+++ +E TEA +G ++ K+ L+ D AG+ I
    Sbjct: 3 VGDIGGGTKIKAALVSDAGEIIISVQECPTEEAQGPEEVNMKMSLTEKVTDHQPFAGIGI 62

50    Query: 65 SSAGMVNPDEGEIFFYAGPQIPNYAGTQFKKEIEETFGLPCVEVDVNCAGLAEAISGSAK 124
            + G ++ EG I + P +P + +E F P +++ND N A LAEA+ GS +
    Sbjct: 63 GAPGPLSSTEGTIL-SPPNLPGWHDHIHLVDRFQEQQCPVQLDNDANVAALAEALLGSGQ 121

55    Query: 125 DYPVALCLTIGTGIGGCLLFNSQVFHGSSHSACEVG-----YLHLSDGQFQDILAS 174
            + LTI TGIGG + + + HG+S A E+G + +L+ G + LAS
    Sbjct: 122 GFTSVFYLTISTGIGGGVLDGSIVHGASDYAGEIGNMIVQPNGYQHANLNPGSLEGILAS 181

    Query: 175 TTALVQEVVVLAYGDDISQWDGRRIFEQAKAGDAICIAISKQVDYLGGQIANICYVVNP 234
            TA+ + +G + R +F+Q + GD + +DYL GIANI + +NP+
    Sbjct: 182 GTAIGRMARERFG---VEGGTREVFDQJIRRGHDHMQRQLVEAMDYLAIGIANIAHTINPD 238
```

Query: 235 VVVLGGGIMAQKDYLADKLKTALDSYLVSSLAKKTQLKFA SHGNNA GILGA 285
 V VLGGG+M D + +K + YL LA+ T + A G ++G+LGA
 Sbjct: 239 VFVLGGGVVNADDLILPIVKEKVSRYLYPGLAQSTTIVKAKLGGDSGV LGA 289

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 192/292 (65%), Positives = 237/292 (80%)

Query: 1 MTRTVAIDIGGTMIKHGIVDNLGCIVEASELATEAYKGGPGILOKVCQIIDNYLAEGSID 60
 M +AIDIGGT IK+G++ G ++E E+ATEAYKGGP IL+KV ++ Y + +
 Sbjct: 1 MKHYLAIDIGGTAIKYGLISETGDLLEKEEMATEAYKGGPSILEVKVGLVKTYQDQMDLA 60

Query: 61 GIAISSAGMVDPDEGCIFYSGPQIPNYAGTQFKKVLEDTYQVRTEIENDVNCAGLAEAVS 120
 G+AISSAGMV+PDEG IFY+GPQIPNYAGTQFKK +E+T+ + E+ENDVNCAGLAEAS
 Sbjct: 61 GVAIASSAGMVNPDEGEIIFYAGPQIPNYAGTQFKKEETFGLPCEVENDVNCAGLAEAS 120

Query: 121 GSAKDSSIALCLTIGTGIGGCLIIDKTVFHGSNSACEVGYMHLSDGDFQDLASTTALIA 180
 GSAKD +ALCLTIGTGIGGCL+ + VFHG S+SACEVGY+HLSDG FQDLASTTAL+
 Sbjct: 121 GSAKDYPVALCLTIGTGIGGCLLFNSQVFHGSNSACEVGYLHLSDGQFQDLASTTALVQ 180

Query: 181 DVAKAHGDEISRWDGRRIFQEAKKGNEKCIAISIDRMINYLQGGIANMVYVVNPKEVVLGG 240
 +V A+GD+IS+WDGRRIF++AK G+ CIA+I + ++YLQGGIAN+ YVNP VVLGG
 Sbjct: 181 EVVLAYGDDISQWDGRRIFEQAKAGDAICIAISKQVDYLQGGIANICYVVNPNVVVLGG 240

Query: 241 GIMAQKDYLQDKLSESLKRNLVTSLAEKTAIVFAQHENQAGMLGAYYHFKNR 292
 GIMAQKDYL DKL +L LV+SIA+KT + FA H N AG+LGAYYHFK +
 Sbjct: 241 GIMAQKDYLADKLKTALDSYLVSSLAKKTQLKFA SHGNNA GILGAYYHFKQK 292

SEQ ID 922 (GBS331) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 2; MW 35.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 3; MW 61kDa).

The GBS331-GST fusion product was purified (Figure 209, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 309), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for 35 vaccines or diagnostics.

Example 288

A DNA sequence (GBSx0315) was identified in *S.agalactiae* <SEQ ID 925> which encodes the amino acid sequence <SEQ ID 926>. This protein is predicted to be a acylneuraminate lyase (nanA). Analysis of this protein sequence reveals the following:

40 Possible site: 18
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0894 (Affirmative) < succ>
 45 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA69950 GB:Y08695 putative acylneuraminate lyase [Clostridium tertium]

Identities = 162/225 (72%), Positives = 191/225 (84%)

Query: 1 MKDLQKYQGIIPAFYACYDDKGDICPERVKALTNYFIDKGVQGLYVNGSSGECIYQS VAD 60
 M++L+KY+GIIPAFYACYDD+G I PER + T Y IDKGV+GLYV GSSGECIYQS +

55 Sbjct: 1 MRNLEKYKGIIPAFYACYDDKGICPERTQMFTQYLIDKGVKG GLYVCGSSGECIYQS VAD 60

-365-

Query: 61 RKLVLENVMSVAKGKLTIVIAHVACNNTKDSVELAMHAEAIGVDAIAAIPPIYFRLPEYAI 120
 RK+ LENVM VAKGK+T+IAHV CNNT+DS ELA HAE+IGVDAIA+IPPIYF LP+Y+I
 Sbjct: 61 RKITLENVMVKAVGKITIIAHVGNCNNTRDSEELAHEASIGVDAIASIPPIYFHLPEYI 120

5

Query: 121 ADYWNTISQAAPQTDIYNIPIQLAGVALTSIDLRYKMLQNPQVIGVKNSSMPVQDIQNFV 180
 A+YWN IS AAP TDFIIYNIPQLAGV L +LY+ML+NP+VIGVKNSSMPVQDIQ F
 Sbjct: 121 AEYWNNDISNAAPNTDFIYNIPIQLAGVGLGINLYKQMLKNPRVIGVKNSSMPVQDIQMFK 180

10

Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAGIGGTYGVMPPELYLTLNQ 225
 I G+ +VFNGPDEQF+ GR+MGA GIGGY VMPEL+L ++
 Sbjct: 181 DISGDESUVFNGPDEQFVAGRIMGADGGGIGGTYAVMPPELFLAADK 225

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 927> which encodes the amino acid
 15 sequence <SEQ ID 928>. Analysis of this protein sequence reveals the following:

Possible site: 15
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.0981(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

25 Identities = 238/304 (78%), Positives = 263/304 (86%)

Query: 1 MKDLQKYQGIIPAFYACYDDKGDICPERVKALTNYFIDKGVQGLYVNGSSGECIYQSVAD 60
 M DL KYQGIIPAFYACYDD+G+I PERV+ALT Y+IDKGVQGLY+NGSSGECIYQSV D
 Sbjct: 1 MTDLTKYQGIIPAFYACYDDQGNISPERVRALTQYYIDKGVQGLYINGSSGECIYQSVFD 60

30 Query: 61 RKLVLENVMSVAKGKLTIVIAHVACNNTKDSVELAMHAEAIGVDAIAAIPPIYFRLPEYAI 120
 R+LVLENVM+VAKGKLT+I HVACNNTKDS+ELA H+E +GVDAIAIPPIYFRLPEYA+
 Sbjct: 61 RQLVLENVMAVAKGKLTIIINHVACNNTKDSIELAAHSERLGVDAAIPPIYFRLPEYAV 120

35 Query: 121 ADYWNTISQAAPQTDIYNIPIQLAGVALTSIDLRYKMLQNPQVIGVKNSSMPVQDIQNFV 180
 ADYWN IS AAP TDFIIYNIPQLAGVALT LY+ ML N +VIGVKNSSMPVQDIQ F
 Sbjct: 121 ADYWNNAISSAAPHTDFIYNIPIQLAGVALTPSLYKTMANKRVIDVKNSSMPVQDIQIFC 180

40 Query: 181 AIGGENHIVFNGPDEQFLGGRLMGAAAGIGGTYGVMPPELYLTLNQLIVDKDLEKARELQF 240
 AIGG++HIVFNGPDEQFLGGRLMGAAAGIGGTYG MPEL+L LNQLI DKDLEKA+ LQ+
 Sbjct: 181 AIGGDDHIVFNGPDEQFLGGRLMGAAAGIGGTYGAMPFLRLNQLIADKDLEKAKALQY 240

45 Query: 241 TINDIITKLCSGHGNMYAVIKAVLEINEQLTIGSVRLPLASVTEEDKPIIKEAAEMIRHA 300
 TIN+II L S HGNMY VIK VL INE L IGSVR PLA + EED+ I + AA +I A
 Sbjct: 241 TINEIIIGVLVSAHGNMYGVVIKEVLRINEGLDIGSVRSPLAELVEEDRVICQRAAALINQA 300

50 Query: 301 KKQF 304
 K+ F
 Sbjct: 301 KETF 304

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 289

A DNA sequence (GBSx0317) was identified in *S.agalactiae* <SEQ ID 929> which encodes the amino acid
 55 sequence <SEQ ID 930>. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -9.45	Transmembrane	82 - 98	(79 - 111)
INTEGRAL	Likelihood = -6.85	Transmembrane	24 - 40	(21 - 52)
INTEGRAL	Likelihood = -5.26	Transmembrane	180 - 196	(172 - 200)

-366-

INTEGRAL Likelihood = -5.10 Transmembrane 160 - 176 (158 - 179)
 INTEGRAL Likelihood = -4.35 Transmembrane 110 - 126 (106 - 130)

----- Final Results -----

5 bacterial membrane --- Certainty=0.4779(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:BAB05827 GB:AP001514 unknown conserved protein in *B. subtilis*
 [Bacillus halodurans]
 Identities = 40/148 (27%), Positives = 74/148 (49%), Gaps = 4/148 (2%)

15 Query: 14 VNNPMQGCNVVFDLALLNLLFMI-TCLPLVTIG--AAKISLYRTLWQKLEGD-QTNLLI 69
 +++ F Q C+ ++ LA +NLL++ T L LV +G A +++ L + G+ +
 Sbjct: 6 MSSRFYQTCDWIWKLAYINNLLWLSGTLLGLVVLGFLPATTAMFTVLRKWFTGNPDVAITR 65

20 Query: 70 LYIKHLKKEWFQGMILLGLVVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFLLLTVMTS 129
 + + K E+ + LLG V L ++ F+ L G+ L + YAFL+L ++T
 Sbjct: 66 TFFQAYKNEFLKINLLGAVLLLGAAYILYFNMYMLGTVEGTVHMLVSLGWYAFLILYIITL 125

25 Query: 130 IYLFPMAARYEMSSLDTVKKSFIMACLN 157
 Y+ P Y + L +K + I+ +N
 Sbjct: 126 FYIIPAYVHYNLKLFQYIKTALIIIGFVN 153

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 931> which encodes the amino acid sequence <SEQ ID 932>. Analysis of this protein sequence reveals the following:

Possible site: 24
 >>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -14.86 Transmembrane 117 - 133 (108 - 139)
 INTEGRAL Likelihood = -7.48 Transmembrane 30 - 46 (21 - 54)
 INTEGRAL Likelihood = -6.90 Transmembrane 88 - 104 (83 - 105)
 INTEGRAL Likelihood = -6.26 Transmembrane 165 - 181 (151 - 187)
 INTEGRAL Likelihood = -5.89 Transmembrane 189 - 205 (182 - 207)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.6944(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:BAB05582 GB:AP001513 unknown conserved protein in bacilli
 [Bacillus halodurans]
 Identities = 59/194 (30%), Positives = 93/194 (47%), Gaps = 11/194 (5%)

45 Query: 17 SKWMRASAALFDLIVFNLLFVL-SCLPLLTIGV--AKMALLYASLLDWREGQVS-QLVTTY 72
 +K M+ + L+ NLL++L S + + +GV A +L+A W + + L TY
 Sbjct: 8 TKIMKLFEWIMRLVYLNLWLLFSFIGGIILGVMPATASLFAVFRKWYQKEDDFPLFQTY 67

50 Query: 73 SSHFKYYFKSGRLRLGLIELGIMTICLDDLFLIRNQSGLVFQGFKVLCVAVLFLVVLFLY 132
 + FK FK +GL + I I LD+ L+ S + Q + A+ F+ ++ LY
 Sbjct: 68 LNEFKRSFKIANIVGLTLVLIGGILYLDVLLLLGTSHWIGQLLLGMVGALSFTIYLVTLLY 127

55 Query: 133 AYPQAVKRDLSSLSTLFLKRSFLLAGLFFPWSFAFLAFICLTFSQL---SLLTLFGGVS 188
 +P V DLS FK SFLL G+ P+ L I L++ +L LL LF S
 Sbjct: 128 IFPTLVHFDSLKYQYFKHSFLL-GVLQPFR-TLLLMTLSALLFLTFPILLPLF-AAS 184

Query: 189 LLAIIGISSLTYLY 202
 +A + + S + Y
 Sbjct: 185 FMAALTMWSFLFGY 198

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 117/210 (55%)

-367-

	Query: 3	KANQLIAAIFDVNNPFMCGNCVVFDLALLNLLFMITCLPLVTIGAAKISLYRTLWQKLEG	62
		K L+ ++F +++ +M+ +FDL + NLLF+++CLPL+TIG AK++LY +L EG	
5	Sbjct: 4	KKQGLLHSLFKLDKWMRASAALFDLLVFNLFLVSLCLPLLTIGVAKMALYASLLDWREG	63
	Query: 63	DQTNLILYIKHLKKEWFQGMILLGLVELSILVVIIFDLTILHYQIGFIVSFLKITCYAFL	122
		+ L+ Y H K + G+ LGL+EL I+ + + DL ++ Q G + K+ C A L	
	Sbjct: 64	QVSQSLVTTYSSHFKYYFKSGRLRLGIELGIMTICLDDLFIRNQSGLVFQGFVKLCAVL	123
10	Query: 123	LLTVMTTSIYLFPMAARYEMLSLLDTVKKSFIMACLNLKWTGVLMFLLIMTWFIMVQSSLLF	182
		L V+ +Y +P A + ++SL K+SF++A L W+ + + +T F + S L	
	Sbjct: 124	FLVVILFLYAYPQAVKRDLSSLTLFKRSFLLAGLFFPWSFAFLAFICLTIFSLSQLSLLTL	183
15	Query: 183	MLTVSATIFIFAYTAFAYFKIIILQKQFAYF	212
		VS + I ++ Y +II++ F	
	Sbjct: 184	FGGVSSLATIIGISSLTYLYLIIMESLLRRF	213

A related GBS gene <SEQ ID 8535> and protein <SEQ ID 8536> were also identified. Analysis of this protein sequence reveals the following:

```

20      Lipop: Possible site: -1    Crend: 2
        McG: Discrim Score:      3.27
        GvH: Signal Score (-7.5): -4.23
              Possible site: 46
25      >>> Seems to have an uncleavable N-term signal seq
        ALOM program count: 5 value: -9.45 threshold: 0.0
              INTEGRAL Likelihood = -9.45 Transmembrane 82 - 98 ( 79 - 111)
              INTEGRAL Likelihood = -6.85 Transmembrane 24 - 40 ( 21 - 52)
              INTEGRAL Likelihood = -5.26 Transmembrane 180 - 196 ( 172 - 200)
              INTEGRAL Likelihood = -5.10 Transmembrane 160 - 176 ( 158 - 179)
30      INTEGRAL Likelihood = -4.35 Transmembrane 110 - 126 ( 106 - 130)
              PERIPHERAL Likelihood = 5.89      142
        modified ALOM score: 2.39

```

35 *** Reasoning Step: 3
----- Final Results -----
 bacterial membrane --- Certainty=0.4779 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

45 ORF00072(364 - 828 of 1260)
EGAD|108353|BS3003(14 - 171 of 222) hypothetical protein {Bacillus subtilis}
OMNI|NT01BS3507 conserved hypothetical protein GP|2635493|emb|CAB14987.1||z99119 similar to
hypothetical proteins from B. subtilis {Bacillus subtilis}
GP|2293197|gb|AAC00275.1|AF008220 YteU {Bacillus subtilis} PIR|D69991|D69991 conserved
hypothetical protein yteU - Bacillus subtilis
%Match = 5.9
%Identity = 26.6 %Similarity = 50.6
Matches = 42 Mismatches = 74 Conservative Sub.s = 38

50

-368-

```

::      || |::::| |::| :: | : :|::
VHFDWKRLYVKFSLLLKVAYLQYLTMLALTVALFFLLAYLPGIVPFFSVSLISYCHMRIVYAVLLKVEQHGGEPKRS
    150      160      170      180      190      200      210

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 290

A DNA sequence (GBSx0318) was identified in *S.agalactiae* <SEQ ID 933> which encodes the amino acid sequence <SEQ ID 934>. Analysis of this protein sequence reveals the following:

```

10 Possible site: 51
     >>> Seems to have no N-terminal signal sequence

     ----- Final Results -----
15      bacterial cytoplasm --- Certainty=0.1827(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:AAC44392 GB:U43526 ORF-1 [Streptococcus pneumoniae]
     Identities = 48/151 (31%), Positives = 66/151 (42%), Gaps = 5/151 (3%)

     Query: 1 MIYDHLLNLTHYKDINPNLDLAIDYLLSHDLRNLDIGTYHISPEVIIMVQSNQLSES-FD 59
             MI + L Y +NP+ ID+L L NL G+ I + L++
     Sbjct: 1 MIITKISRLLGTYVGVNPHFATLIDFLEKTGLENLTEGSIAIDGNRLFGNCFTYLADGQAG 60
25     Query: 60 HIFEYHKYLDIHYVIEGHIVEIKLGKGDKVEV-EEY--LGDIGFIKCEETSFDLRDNYI 116
             FE H+KYLDIH V+E E + + + V V +EY DI E LR
     Sbjct: 61 AFFETHQKYLDIHLVLENNEAMAVTSPENVTQEYDEEKDIELYTGKVEQLVHLRAGEC 120

     Query: 117 AFFFPEEAHQPNMGSLGNVVKGVLKVLMA 147
             FPE+ HQP + VKV V KV ++
     Sbjct: 121 LITFPEDLHQPK-VRINDEPVKKVVFKAIS 150

```

No corresponding DNA sequence was identified in *S.pyogenes*.

- 35 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 291

A DNA sequence (GBSx0319) was identified in *S.agalactiae* <SEQ ID 935> which encodes the amino acid sequence <SEQ ID 936>. This protein is predicted to be sugar ABC transporter, permease protein (araQ).

- 40 Analysis of this protein sequence reveals the following:

```

Possible site: 35
>>> Seems to have a cleavable N-term signal seq.

     INTEGRAL Likelihood = -7.38 Transmembrane 245 - 261 ( 239 - 265)
     INTEGRAL Likelihood = -3.72 Transmembrane 140 - 156 ( 139 - 158)
45     INTEGRAL Likelihood = -3.61 Transmembrane 76 - 92 ( 71 - 94)
     INTEGRAL Likelihood = -2.81 Transmembrane 112 - 128 ( 107 - 128)
     INTEGRAL Likelihood = -1.59 Transmembrane 188 - 204 ( 186 - 204)

     ----- Final Results -----
50      bacterial membrane --- Certainty=0.3951(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

-369-

>GP:AAD35515 GB:AE001721 sugar ABC transporter, permease protein
 [Thermotoga maritima]
 Identities = 94/262 (35%), Positives = 158/262 (59%), Gaps = 1/262 (0%)

5 Query: 15 LILCLLTVLFIFPFYWIMTGAFKSQPD TIIIPQQWPKAPTLENFKALT VQNPALRWLN 74
 + + + V+F+ P ++ + +FK + PP +PK P+LE + + + L +L N
 Sbjct: 9 IFIGVPMVVFMPLPVYAVVSSFKPMSEIYSYPPTIFPPKKPSLEGYINVIEKEYDLLTYLRN 68

10 Query: 75 SVFISIMTMFLVCCTSSMAGYVIAKRFYQGQKILFSLFIAAMALPKQVVLVPLVRIINFM 134
 ++F++ + + S M GY IAK +F+G + + S+F M + QV++VPL +I +
 Sbjct: 69 TLFVATVATVATVITVIVSVMTCGYGLAKGKFWGIRPVNSMFTMTMFVSAQVIMVPLFVVIRSL 128

15 Query: 135 GIHD TLWA VILPLVGWPGFVFLMKQFSENIPTELLESAKIDGC GEIRTFINVA FPIVKPG 194
 G+ ++LW +I+P V P G+F+ Q+ ++IP ELLESAKIDG E + F + FP+ KP
 Sbjct: 129 GLINSLWGLIIPAVYPTGMFMAVQYMKDIPDELLESAKIDGANEWQIFWRIVFPLSKPL 188

20 Query: 195 FAALAI FT FINTWNDYFMQLVMLTSRNNLTISLGVATM QAEM-ATNYGLIMAGAALAAVP 253
 AALAIF+F WND+ L+++ RN T+ L +AT+Q E + I+A + L +P
 Sbjct: 189 VA ALAIF SFTW RWNDFVLP PLLVVNRNLYTLQLALATIQEEYGGAEWNTILAFSTLT IIP 248

Query: 254 IVTVFLVFKQSFTQGITMGAVK 275
 + +FL+ FQ+ F +GI G +K
 Sbjct: 249 TLII FLLFQRLFMKGIMAGGLK 270

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 937> which encodes the amino acid sequence <SEQ ID 938>. Analysis of this protein sequence reveals the following:

Possible site: 40
 >>> Seems to have a cleavable N-term signal seq.
 30 INTEGRAL Likelihood = -6.37 Transmembrane 245 - 261 (240 - 265)
 INTEGRAL Likelihood = -5.15 Transmembrane 140 - 156 (139 - 158)
 INTEGRAL Likelihood = -2.97 Transmembrane 111 - 127 (107 - 128)
 INTEGRAL Likelihood = -2.87 Transmembrane 76 - 92 (75 - 93)
 INTEGRAL Likelihood = -1.59 Transmembrane 188 - 204 (186 - 204)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.3548(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has homology with the following sequences in the databases:

>GP:CAB59597 GB:AL132662 probable sugar transport inner membrane
 protein [Streptomyces coelicolor A3(2)]
 Identities = 88/262 (33%), Positives = 147/262 (55%)

45 Query: 15 VMLCVL TILFIFPFYWIMTGAFKAQADTMI PQQWPKAPTLENFKALV VQNPALKWLWN 74
 ++L L ++F P W++ + +A+ PP WP + ++ ++ +W N
 Sbjct: 38 LLLAPLALVFAVPLVWLVLSSVMSNAEINRFP PALWPSGIDLGGYRYVLGNAMFPRWFVN 97

50 Query: 75 SVFISVATMFLVCGTSSLAGYALAKRFYQGQRLLSIFIAAMALPKQVVLVPLVRIVNFM 134
 S+ +S T+ SLAGYA A+ RF G R+L + +A MA+P Q+ ++P ++ +
 Sbjct: 98 SLIVSAVTVAANLVFGSLAGYAFARMRFAGSRVLMGLMLATMAVPFQLTMIPTFLVMKKL 157

Query: 135 GIHD TLAA VILPLVGWPGFVFLMKQFSENIPTELLESAKIDGC GEIRTFNVA FPIVKPG 194
 G+ DTL A+I+P + PF VFL++QF ++P EL E+A IDGC +R + + P+ +P
 Sbjct: 158 GLID TL GALIVPSLVT PFAVFLLRQFFLSPRELEEA WIDGCSRLRVLW RIVLPLSRPA 217

Query: 195 FAALAI FT FINTWNDYFMQLVMLTSRENLTISLGVATM QAEMATNYGLIMAGAAMAAVP 254
 A +A+ TF+ TWND L+ + T+ LG+ T Q+ T+ +MAG + +P+
 Sbjct: 218 LATVAVL TLTTWNDLTWPLIAINHDTQYTLQLGLTFQGQHHTQWA AVMAGNVITVLPV 277

60 Query: 255 VTVFLVFKQSFTQGITMGAVK 276
 + FL QK+F Q IT +KG
 Sbjct: 278 LLAFLGAQKTFIQSITSSGLKG 299

65 An alignment of the GAS and GBS proteins is shown below:

-370-

Identities = 245/276 (88%), Positives = 262/276 (94%)

Query: 1 MKKKTF SAYNFLT ALILCLLTVLFIFPFYWIMTGAFKSQPD TIIIPPQWWPKAPTLENFK 60
 5 Sbjct: 1 MTKKKLTAS DILTTVMLCVL TILFIFPFYWIMTGAFKAQADT I MIPPQWWPKAPTENFK 60

Query: 61 ALTVQNPALRWLWNSVFISIMMFLVCCTSSMAGYVLAKKR FYGQKILFSLFIAAMALPK 120
 10 Sbjct: 61 ALVVQNPALKWLWNSVFISVATMFLVC GTSSLAGYALAKKR FYGQRLLFSIFTIAAMALPK 120

Query: 121 QVVLVPLVRI INF NFMGIHD TLWAVI LPLVG WFGVFLMKQFSENIPTELLESAKIDGC GEI 180
 15 Sbjct: 121 QVVLVPLVRI VNF NFMGIHD TLAAV I LPLVG WFGVFLMKQFSENIPTELLESAKIDGC GEI 180

Query: 181 RTF INVA FPIVKPGFAA LAIFT FINTWNDY FMQLVMLTSR NNLT ISLG VATM QAE MATN Y 240
 Sbjct: 181 RTFFNVA FPIVKPGFAA LAIFT FINTWNDY FMQLVMLTSR ENLT ISLG VATM QAE MATN Y 240

Query: 241 GLIMAGA AALAA VPIV TVFLV FQKSFT QG GIT MGAV KG 276
 20 Sbjct: 241 GLIMAGA AAMAA VPIV TVFLV FQKSFT QG GIT MGAV KG 276

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 292

A DNA sequence (GBSx0320) was identified in *S.agalactiae* <SEQ ID 939> which encodes the amino acid sequence <SEQ ID 940>. Analysis of this protein sequence reveals the following:

Possible site: 31
 30 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -10.83 Transmembrane 74 - 90 (64 - 96)
 INTEGRAL Likelihood = -6.37 Transmembrane 108 - 124 (107 - 126)
 INTEGRAL Likelihood = -5.84 Transmembrane 270 - 286 (265 - 290)
 INTEGRAL Likelihood = -5.20 Transmembrane 161 - 177 (156 - 182)
 INTEGRAL Likelihood = -0.16 Transmembrane 219 - 235 (219 - 235)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.5331(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40

The protein has homology with the following sequences in the GENPEPT database:

>GP: BAB05584 GB: AP001513 sugar transport system (permease) (binding protein dependent transporter) [Bacillus halodurans]
 45 Identities = 106/289 (36%), Positives = 168/289 (57%), Gaps = 6/289 (2%)

Query: 9 RETMIA YAFLA PILLFFLIFV FAPMVGMGFV TSFFNYSM-TQFTFIGL ANYNRMF-HDSIF 66
 +E Y F+AP ++ F IF PM+ SF ++ + + + G NY R+F D +F
 Sbjct: 25 KEYFWGYLFIA PPIIGFAI FALGPMLYSIYVSFTDFDLYNEPVWTGADNYYRLFVTDDL F 84

50 Query: 67 MKSLINTVIIVIGSVPVVVFFSLFVAANTYEK NVFSRSFYRCVFFLPVVTGSVAVT VVWK 126
 K++ NT +G +P+ + SL +A +K V + +R FFLP V+ VA+T++W+
 Sbjct: 85 RKT VFNTFYAALG- IPIGM AVSLGIAVALNQK-VKGIALFR TAFFLPAVSSVVAITLLWR 142

55 Query: 127 WIYDPMSGILNYILKSGH VIEQNISWLGDKH WALLA II ILLT SVGQPII LYIAAMGNI 186
 WI++ G+LN +L +V WL D+ WA+ A+II + +G +ILY+AA+ +
 Sbjct: 143 WIFNA DFG LLNIMLN--YVG I HGP GWL SDE KWAMPAMII QGVWGGLGINMILYLAALQGV 200

Query: 187 DNSLCEAARVDGANEMQVFWQIKWPSLLPTTLYIAVITTINSFQCFALIQLLTSGGP NYS 246
 + +L EAA +DG N Q F I PS+ PTT +I +TI + Q F ++T GGP NYS
 60 Sbjct: 201 NPAL YEAADIDGGNAWQKFTHITVPSISPTFFFILITSTIGALQDFQR FMIMTEGGP NYS 260

Query: 247 TSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIALISFAQFKILGNDVEY 295

-371-

T+T++YYL+ AF+ E GYA+ M L ++I +I+ FK+ V Y
 Sbjct: 261 TTTVVVYLFN AFRYMEMGYASAMAWVLGIIILITIINFKLAKKWHY 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 941> which encodes the amino acid
 5 sequence <SEQ ID 942>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 10 INTEGRAL Likelihood = -12.74 Transmembrane 55 - 71 (44 - 78)
 INTEGRAL Likelihood = -10.83 Transmembrane 109 - 125 (98 - 130)
 INTEGRAL Likelihood = -6.21 Transmembrane 304 - 320 (299 - 324)
 INTEGRAL Likelihood = -6.00 Transmembrane 142 - 158 (141 - 160)
 INTEGRAL Likelihood = -5.04 Transmembrane 196 - 212 (190 - 216)
 INTEGRAL Likelihood = -0.16 Transmembrane 253 - 269 (253 - 269)
 15 ----- Final Results -----
 bacterial membrane --- Certainty=0.6095 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20 The protein has homology with the following sequences in the databases:

>GP:BAB05584 GB:AP001513 sugar transport system (permease) (binding
 protein dependent transporter) [Bacillus halodurans]
 Identities = 113/310 (36%), Positives = 176/310 (56%), Gaps = 9/310 (2%)
 25 Query: 25 KVEQKKEVFQVNVLKMR---ETLISYAF LAPVLVFFVI FVLIPMIMGFVTSFFNYSM- 80
 +VE +E K K R E Y F+AP ++ F IF L PM+ SF ++ +
 Sbjct: 4 EVETPRETKTTKARKQKRRLNKEYFWGYLFIAPPIIGFAIFALGPMLYSIYVSFTDFDLY 63
 30 Query: 81 TEFTFVGFANYARMF-QDPIFMKSLINTLIIIVIGSVPVVFFSLFVAAKTYDKNVVARSF 139
 E + G NY R+F D +F K++ NT +G +P+ + SL +A KV +
 Sbjct: 64 NEPVWTGADNYYRLFVTDDLFRKTVFNTFYAALG- IPIGMASVLSGIAVALNQK-VKGIAL 121
 Query: 140 YRAVEFLPVVTGSAVITVVKWYIDPMMSGILNYVLKYAHVIEQNISWLGDKH WALLAIIV 199
 +R FF LP V+ VA+T++W+WI++ G+LN +L Y + WL D+ WA+ A+I+
 35 Sbjct: 122 FRTAFFLP AVSSVVAITLLWRWIFNADFGLLNIMLNIVGII--HGPGLSDEKWAMPAMII 179
 Query: 200 ILLTTSVGQPIIILYIAAMGNIDNSLVEAARVDGATEFQVFWNIKWPSSLPTTLYIAVITT 259
 + +G +ILY+AA+ ++ +L EAA +DG +Q F +I PS+ PTT +I + +T
 Sbjct: 180 QGVWGGLGINMILYLAQGVNPALYEAADIDGGNAWQKFIHITVPSISPTFFFILITST 239
 40 Query: 260 INSFQCFALIQLLTSGGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQ 319
 I + Q F ++T GGPNYST+T++YYL+ AF+ E GYA+ M L ++I III+
 Sbjct: 240 IGALQDFQRFMIMTEGGPNYSTT VVYLFN AFRYMEMGYASAMAWVLGIIILITIIN 299
 45 Query: 320 FKILGNDVEY 329
 FK+ V Y
 Sbjct: 300 FKLAKKWHY 309

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 263/295 (89%), Positives = 278/295 (94%)
 Query: 1 MRTNKLKMRETMIA Y AFLAPILLFFLIFV FAPMV MG FV TSFFNYSMTQFTFIGLANYNRM 60
 + NKLKMRET+I+YAFLAP+L+FF+IFV PM+MGFV TSFFNYSMT+FTF+G ANY RM
 Sbjct: 35 VNVNKLKMRETLISYAF LAPVLVFFVI FVLIPMIMGFVTSFFNYSMT EFTFVGFANYARM 94
 55 Query: 61 FHDSIFMKSLINTVIIVIGSVPVVFFSLFVAANTYEKNFSRSFYRCVFLPVVTGSVA 120
 F D IFMKSLINT+IIVIGSVPVVFFSLFVA TY+KNV +RSFYR VF FLPVVTGSVA
 Sbjct: 95 FQDPIFMKSLINTLIIIVIGSVPVVFFSLFVA AKTYDKNVVARSFYRAVFFLPVVTGSVA 154
 60 Query: 121 VTVVWKWYIDPMMSGILNYILKSGH VIEQNISWLGDKH WALLAI IIII LTT SVGQPIIILYI 180
 VTVVWKWYIDPMMSGILNY+LK HVIEQNISWLGDKH WALLAI II+III LTT SVGQPIIILYI
 Sbjct: 155 VTVVWKWYIDPMMSGILNYVLKYAHVIEQNISWLGDKH WALLAI IV LTT SVGQPIIILYI 214
 Query: 181 AAMGNIDNSLCEAARVDGANEMQVFWQIKWPSSLPTTLYIAVITTINSFQCFALIQLLTS 240

-372-

```

AAMGNIDNSL EAARVDGA E QVFW IKWPSLLPTTLYIAVITTINSFQCFALIQLLTS
Sbjct: 215 AAMGNIDNSLVEAARVDGATEFQVFVNKWPSSLPTTLYIAVITTINSFQCFALIQLLTS 274

Query: 241 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIALISFAQFKILGNDVEY 295
      GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIA+ISFAQFKILGNDVEY
Sbjct: 275 GGPNYSTSTLMYYLYEKAFKLSEYGYANTMGVFLAVMIAIISFAQFKILGNDVEY 329

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 293

A DNA sequence (GBSx0321) was identified in *S.agalactiae* <SEQ ID 943> which encodes the amino acid sequence <SEQ ID 944>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have a cleavable N-term signal seq.

15 ----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20 The protein has homology with the following sequences in the GENPEPT database:

```

```

>GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
  Identities = 54/187 (28%), Positives = 90/187 (47%), Gaps = 14/187 (7%)

```

```

25 Query: 19 MFACVDSSQSVMMAEKD-KVEITWWAFPTFTQEKAQDGVTGYEKKVIKAFEEKKNPNIKVK 77
      MF+ + + ++D + I WW + D Y KVI+ +EKKNP++ ++
Sbjct: 1 MFSGCSAGEEASGKEDVTIRIAWWG-----GQPRHD-----YTTKVIELYEKKNPHVHIE 51

```

```

30 Query: 78 LETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNGKLAIDLNDLFTDQFIKDVN-- 135
      E ++ +K+ AG PDV+ + QYGK +L DL D I DV+
Sbjct: 52 AEFANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTPYTKDGTI-DVSSI 110

```

```

35 Query: 136 NKNIIQASKSGDKAYMYPISSAPFYMAFNKMLKDAGVLKLVKEGWTTSDFEKVLKALKN 195
      ++N++ K +K Y + + + N+ +LK AGV + +E WT D+EK+ L+
Sbjct: 111 DENMLSGGKIDNKLYGFTLGVNVLSVIANEDLLKKAGV-SINQENWTWEDYEKLAYDLOE 169

```

```

40 Query: 196 KGTYPGS 202
      K GS
Sbjct: 170 KAGVYGS 176

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 945> which encodes the amino acid sequence <SEQ ID 946>. Analysis of this protein sequence reveals the following:

Possible site: 20

>>> May be a lipoprotein

```

45 ----- Final Results -----
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

The protein has homology with the following sequences in the databases:

```

!GB:Z99107 similar to sugar-binding protein [Bacillu... 82 2e-14

```

```

55 >GP:CAB12516 GB:Z99107 similar to sugar-binding protein [Bacillus subtilis]
  Identities = 105/446 (23%), Positives = 176/446 (38%), Gaps = 71/446 (15%)

```

```

Query: 24 GKSQKEAGASKSDTAKTEITWWAFPVFTQEKAEDGVGTYEKKLIAAFEKANPEIKVKLET 83
      G S E + K + I WW + D Y K+I +EK NP + ++ E

```

-373-

5 Sbjct: 4 GCSAGEEASGKKEVTLRIAWWG----GQPRHD---YTTKVIELYEKKNPHVHIEAF 54

Query: 84 IDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNGKLADLNLFTEFTKDVN--NDK 141
++ +K+ AG PDV+ + QYKG +L DL +T++ T DV+ ++

5 Sbjct: 55 ANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTP-YTKDGTIDVSSIDEN 113

Query: 142 LIQASKAGDTAYMYPISAPFYMALNKKMLKDAGVLDLVKEGWTDDFEKVLKALKDK-- 199
++ K + Y ++ + N+ +LK AGV + E WT +D+EK+ L++K

10 Sbjct: 114 MLSGGKIDNKLYGFTLGVNVLSVIANEDLLKKAGV-SINQENWTIWEDYEKLAYDLQEKG 172

Query: 200 ----GYNPGSFFANGQGGDQGPRAFFANLYSSHITDDKV-----TKYTT 239
G +P F +G R + + DD++ T T

Sbjct: 173 VYGSNGMHPPDIFFPYIYLRTKGERFYKEDGTGLAYQDDQLFVDYFERQLRLVKAKTSPTP 232

15 Query: 240 DDANSIKAMTKISNWIKDGMLMMNGSQYDGSADIQNPFANGQTSFTILWAPAQPGIQAKLLE 299
D++ IK M +D ++ G SA N++N F A+L +

Sbjct: 233 DESAQIKGM-----EDDFIVKGK---SAITWNYSNQYLG-----ARLTD 269

20 Query: 300 ASKVODYLEIPFPSDDGKPELEYLVNGFAVFNNKDEQKVAASKTFIQLFIADDKEWGPKNVV 359
+ YL P + L + E K A+K FI F +++E + +

Sbjct: 270 SPLSLYLP---PEQMQEKAFTLKPMSMLFSIPKSSEHKKEAAK-FINFFVNNEE-ANQLIK 324

Query: 360 RTGAFPVRTSYGDLYKDKRMEK---IAEWTKFYSPTYNTID----GFAEMRTLWFPMVQ 411
PV D K K E+ I E+ + S + D G AE+ L

25 Sbjct: 325 GERGVPSDKVADAIPKPLNEETNIVEYVETASKNISKADPPEPVGSAEVIKLLKDTSD 384

Query: 412 AVSNGDEKPEDALKAFTEKANKTIKK 437
+ PE A K F +KAN+ +++

30 Sbjct: 385 QILYQKVSPEKAATFRKKANEILER 410

An alignment of the GAS and GBS proteins is shown below:

Identities = 352/438 (80%), Positives = 384/438 (87%), Gaps = 4/438 (0%)

Query: 1 MSIKKSVIGFCLGAAALSMFACVDSSQSVMMAEKD---KVEITWWAFPTFTQEKAQDGVG 57
M++KK LGA+ L + AC SQ A K K EITWWAFT FTQEKA+DGVG

35 Sbjct: 1 MNMKKLASLAMLGASVLGLAACGGKSQKEAGASKSDTAKTEITWWAFPVFTQEKAEDGVG 60

Query: 58 TYEKKVIKAFEKKPNIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG 117
TYEKK+I AFEK NP IKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG

40 Sbjct: 61 TYEKKLIAAFEKANPEIKVKLETIDFTSGPEKITTAIEAGTAPDVLFDAPGRIIQYGKNG 120

Query: 118 KLAIDLNLFTDQFIKDVNNKNIIQASKSGDKAYMYPISAPFYMAFNKKMLKDAGVLKLV 177
KLAIDLNLFT++F KVNNN +IQASK+GD AYMPISAPFYMA NKKMLKDAGVL LV

45 Sbjct: 121 KLAIDLNLFTEEFTKDVNNDKLIQASKAGDTAYMYPISAPFYMAFNKKMLKDAGVL 180

Query: 178 KEGWTTSDFEKVLKALKNGYTPGSFFANGQGGDQGPRAFFANLYSAPITDKEVTKYTTD 237
KEGWTT DFEKVLKALK+KGY PGSSFANGQGGDQGPRAFFANLYS+ ITD +VTKYTTD

Sbjct: 181 KEGWTTDDFEKVLKALKDKGYNPGSFFANGQGGDQGPRAFFANLYSSHITDDKVTKYTTD 240

50 Query: 238 TKNSVKSMKKIVEWIKGYLMNGSQYDGSADIQNPFANGQTAFTILWAPAQPKTQAKLLES 297
NS+K+M KI WIK G +MNGSQYDGSADIQNPFANGQT+FTILWAPAQP QAKLLE+
Sbjct: 241 DANSIKAMTKISNWIKDGMLMMNGSQYDGSADIQNPFANGQTSFTILWAPAQPGIQAKLLEA 300

55 Query: 298 SKVDYLEVPFPSSEDGKPDLEYLVNGFAVFNNKDENKVASKKFITFIADDKKWGPKDVR 357
SKVDYLE+PFPS+DGKP+LEYLVNGFAVFNNKDE KV ASK FI FIADDK+WGPK+V+R

Sbjct: 301 SKVDYLEIPFPSDDGKPELEYLVNGFAVFNNKDEQKVAASKTFIQLFIADDKEWGPKNVVR 360

Query: 358 TGAFPVRTSYGDLYKGDKRMMKISKWTQYYSPYYNTIDGFSEMRTLWFPMVQSVSNGDEK 417
TGAFPVRTS+GDLYK DKRM KI++WT++YSPYYNTIDGF+EMRTLWFPMVQ+VSNGDEK

60 Sbjct: 361 TGAFPVRTSYGDLYK-DKRMKIAEWTKFYSPTYNTIDGFAEMRTLWFPMVQAVSNGDEK 419

Query: 418 PADALKDFTQKANDTIKK 435
P DALK FT+KAN TIKK

Sbjct: 420 PEDALKAFTEKANKTIKK 437

-374-

A related GBS gene <SEQ ID 8537> and protein <SEQ ID 8538> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 4
 5 McG: Discrim Score: 5.05
 GvH: Signal Score (-7.5): 4.69
 Possible site: 31
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 7.69 threshold: 0.0
 PERIPHERAL Likelihood = 7.69 90
 10 modified ALOM score: -2.04

*** Reasoning Step: 3

----- Final Results -----

15 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

20 28.8/48.4% over 409aa

Bacillus subtilis

EGAD|107689| hypothetical protein Insert characterized
 GP|2633010|emb|CAB12516.1||Z99107 similar to sugar-binding protein Insert characterized
 PIR|F69796|F69796 sugar-binding protein homolog yes0 - Insert characterized

25 ORF01146 (355 - 1605 of 1914)

EGAD|107689|BS0697(1 - 410 of 412) hypothetical protein {Bacillus subtilis}GP|2633010|emb|CAB12516.1||Z99107 similar to sugar-bindin g protein {Bacillus subtilis}PIR|F69796|F69796 sugar-binding protein homolog yes0 - Bacillus subtilis
 %Match = 5.4
 %Identity = 28.8 %Similarity = 48.3
 Matches = 69 Mismatches = 116 Conservative Sub.s = 47

30 35 318 348 378 435 465 495 525
 RGIVMSIKSVIGFCLGAAALSMFACVDSSQSVMAAEKD-KVEITWWAFPTFTQEAKADGVGTYEKKVIAKAFEKKNPNIK
 ||: : : :: | : | || | |||: : |||:::
 MFSGCSAGEEASGKKEDVTLRIAWW-----GGQPRHDYTTKVIELYEKKNPVH

10 20 30 40

40 45 555 585 615 645 675 705 732 762
 VKLETIDFTSGPEKITTAIEAGTAPDVLFADPGRRIIQYGKNGKLADLNDLFTDQFIKDVN-NKNIIQASKSGDKAYMYP I
 :: | : : : | || | : || | : || | | : : : | : | : :
 IEAEFANWDDYWKKLAPMSAAGQLPDVIQMDTAYLAQYGKKNQLEDLTPYTKDTIDVSSIDENMLSGGKIDNKLYGF TL
 60 70 80 90 100 110 120

792 822 852 882 912 942 972
 SSAPFYMAFNKKMLKDAGVLKLVKEGWTTSDFEKVLKALKNKGYTPGSFFANGQGGDQGPRAFFANLYSA-----

50 140 150 160 170 180 190 200
 GVNVLSTIANEDLLKKAGV-SINQENWTWEDYEKLAYDLQEK--AGVYGSNGM--HPPDFIPYYLRTKGERFYKEDG

990 1020 1050 1080
 -----PITDKEVTKYTTDTKNSVKSMKKIVIEWIKKGYLMNGSQYDGSA~~~
 | : | | | : : |
 TGLAYQDDQL~~~NIVEYVETASKNISKADPPPEPVGSAEVIKLLKDTSDQILYQKV-----
 350 360 370 380 390

55 60 1515 1545 1575 1605 1635 1665 1695 1725
 FSEMRTLWFPVMQSVSNGDEKPADALKDFTQKANDTIKKAAK*LRRLLFYQSHIGIEEEFLVCLRCKGEYRMRTNKL K
 | | | | : ||: : :
 -----SPEKAAKTFRKKANEILERNN

SEQ ID 944 (GBS16) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 3 (lane 9; MW 49kDa).

The GBS16-His fusion product was purified (Figure 92A; see also Figure 189, lane 9) and used to immunise mice (lane 1 + 2 product; 20 μ g/mouse). The resulting antiserum was used for Western blot (Figure 92B),

5 FACS (Figure 92C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 294

10 A DNA sequence (GBSx0322) was identified in *S.agalactiae* <SEQ ID 947> which encodes the amino acid sequence <SEQ ID 948>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

20 A related GBS nucleic acid sequence <SEQ ID 9459> which encodes amino acid sequence <SEQ ID 9460> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP: AAC66999 GB: AE001166 conserved hypothetical protein [Borrelia
burgdorferi]
25   Identities = 107/225 (47%), Positives = 147/225 (64%), Gaps = 6/225 (2%)

      Query: 12 QIKNGIIVSQALPGEPLYTESGGVMPPLLALAAQEAAGAVGIRANSVRDIKEIQEVTNLPI 71
              +IK G+IVSCQAL EPL+ S +M +ALAA+ GA+GIRAN V DI +I+ +LPI
      Sbjct: 6 KIKRGLIVSQALENEPLH--SSFIMSKMALAAKIGGAIGIRANGVNDISQIKLEVSDLPI 63

30      Query: 72 IGIKKREYPPQEPPFITATMTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQKIKRKYP 131
              IGIKK+ Y + FIT TM E+D+L + + +IALD T R R DG+ + +F + IK+KYP
      Sbjct: 64 IGIKKKNNNCDVFITPTMKEIDELCNEGVDIIALDATFRNRPDGVLDDFFENIKKKYP 123

      Query: 132 EQLLMADISTFEEGKNAFEAGVDFVGTTLSGYTDYSR--QEEGPDIELLNKLCQAGI--D 187
              +Q LMADIS+ +E NA + G DF+GTTL GYT + D L L + +
      Sbjct: 124 KQCLMADISSLDEAINADKLGFDFIGTTLGYTKNTNGLNIADNDNFNLRTLLNSNLKST 183

      Query: 188 VIAEGKIHTPKQANEINHIGVAGIVVGGAITRPKEIAERFISGLS 232
              +I EGKI TP +A + +GV +VVGGAITRP EI ++F+ ++
      Sbjct: 184 LIVEGKIDTPLKAQKCFEMGVDLVVVGGAITRPAEITKKFVEKIN 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 949> which encodes the amino acid sequence <SEQ ID 950>. Analysis of this protein sequence reveals the following:

```
45   Possible site: 44
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -1.49 Transmembrane 175 - 191 ( 175 - 192)

      ----- Final Results -----
50      bacterial membrane --- Certainty=0.1595(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD28762 GB:AF130859 putative N-acetylmannosamine-6-P epimerase
  [Clostridium perfringens]
  Identities = 113/225 (50%), Positives = 148/225 (65%), Gaps = 5/225 (2%)
5
Query: 10 LMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAAAQEAGAVGIRANSVRDIKEIQAID 69
        +++ +KG +IVSCQAL EPL+S IM MA AA++ GA IRA + DI EI+ +T
Sbjct: 1 MLDVVKGNLIVSCQALSDEPLHSSF--IMGRMAIAAKQGAAAIRAQGIDDINEIKEVTK 58
10
Query: 70 LPTIGIICKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKDRHDGLDIASFIRQVKE 129
        LPIIGIIK++Y E +IT TM EVD+L + +I +D TKR R +G +I + +
Sbjct: 59 LPTIGIICKRNYDDSEIYITPTMKEVDELLKTDCMIGLDAKRPNGENIKDLVDAIHA 118
15
Query: 130 KYPNQLLMADISTFDEGLVAHQAGIDFVGTTLSGYTPYSRQEAGPDVALIEALCK-AGIA 188
        K +L MADIST +EG+ A + G D V TTLSGYTPYS+Q D L+E L K I
Sbjct: 119 K--GRLAMADISTLEEGIEAEKLGFDVCSTTLSGYTPYSKQSNSVDFELLEELVKTVKIP 176
Query: 189 VIAEGKIHSPEEAKKINDLGVAGIVVGGAITRKPKEIAERFIEALK 233
        VI EG+I++PEE KK DLG VVGGAITR+I +RF + LK
20
Sbjct: 177 VICEGRINTPEELKKALDLGAYSAVVGGAITRQQITKRFTDILK 221
```

An alignment of the GAS and GBS proteins is shown below:

Identities = 172/227 (75%), Positives = 202/227 (88%)

```
25
Query: 5 SKEAFKKQIKNGIIVSCQALPGEPLYTESGGVMPLLALAAQEAGAVGIRANSVRDIKEIQ 64
        +KE +Q+K GIIVSCQALPGEPLY+E+GG+MPL+A AAQEAGAVGIRANSVRDIKEIQ
Sbjct: 6 TKEKLMEQLKGGIIVSCQALPGEPLYSETGGIMPLMAAAQEAGAVGIRANSVRDIKEIQ 65
30
Query: 65 EVTNLPPIIGIICKREYPPQEPFITATMTEVDQLASLDIAVIALDCTLRERHDGLSVVEFIQ 124
        +T+LPIIGIIK++YPPQEPFITATMTEVDQLA+L+IAVIA+DCT R+RHDGL + FI+
Sbjct: 66 AITDLPIIGIICKDYPPQEPFITATMTEVDQLAALNIAVIAMDCTKDRHDGLDIASFIR 125
35
Query: 125 KIKRKYPEQQLMADISTFEEGKNAFEAGVDFVGTTLSGYTDYSRQEEGPDIELLNLKCQA 184
        ++K KYP QILLMADISTF+EG A +AG+DFVGTTLSGYT YSRQE GPD+ L+ LC+A
Sbjct: 126 QVKEKYPNQLLMADISTFDEGLVAHQAGIDFVGTTLSGYTPYSRQEAGPDVALIEALCKA 185
Query: 185 GIDVIAEGKIHTPKQANEINHIGVAGIVVGGAITRKPKEIAERFISGL 231
        GI VIAEGKIH+B++A +IN +GVAGIVVGGAITRKPKEIAERFI L
40
Sbjct: 186 GIAVIAEGKIHSPEEAKKINDLGVAGIVVGGAITRKPKEIAERFIEALK 232
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 295

A DNA sequence (GBSx0323) was identified in *S.agalactiae* <SEQ ID 951> which encodes the amino acid sequence <SEQ ID 952>. This protein is predicted to be group B streptococcal surface immunogenic protein. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have a cleavable N-term signal seq.

50 ----- Final Results -----
      bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

55 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 953> which encodes the amino acid sequence <SEQ ID 954>. Analysis of this protein sequence reveals the following:

Possible site: 25

-377-

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

5 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 182/437 (41%), Positives = 240/437 (54%), Gaps = 53/437 (12%)

Query: 1 MKMNKKVLLTSTMAASLLSVASVQAQETDTTWTARTVSEVKADLVKQDNKSSYTAKYGDT 60
 M + KK L +++A SL+ +A+ QAQE WT R+V+E+K++LV DN +YTVKYGDT
 Sbjct: 1 MIITKKSLFVTSVALSLVPLATAQAQE----WTPRSVTEIKSELVLVDNVFTYTVKYGDT 56

15 Query: 61 LSVISEAMSIDMNVLAKINNIADINLIYPETTLTVTYDQKSHTATSMKIETPATNAAGQT 120
 LS I+EAM ID++VL IN+IA+I+LI+P+T LT Y+Q AT++ ++ PA++ A +
 Sbjct: 57 LSTIAEAMGIDVHVLGDINHIANIDLIFPDILTANYNQHQ-ATNLTVQAPASSPASVS 115

20 Query: 121 TATVDLKTNQVSADQKVSINTSEGMLTP-EAATTIVSPMKTYSSAPALKSKEVLAQEQA 179
 Q S Q ++ TP + TT + K SS A S E+ +
 Sbjct: 116 HVPSSEPLPQASATSQPTV--PMAPPATPSDVPTPFASAKPDSSVTA--SSELTSSTND 171

25 Query: 180 VSQAAANEQVSPAPVKSITSEVPAAKEEVKPTQTSVSQSTTVSPASVAAETPAPVAKVAP 239
 VS ++E V P A E T V T +S A +A P P +
 Sbjct: 172 VSTELSSESQKQPEVPQEAIVPTPKAAE----TTEVEPKTDISEAPTSANRPVPNESASE 226

30 Query: 240 VRTVAAPRVASVKVTPKVETGASPEHVSAPAVP---VTTSPATDSKLQATEVKSVPA 296
 + AAP + A E SAPA TTS AT + L
 Sbjct: 227 EVSSAAP-----AQAPAEKEETSAPAAQKAVADTTSVATSNGL----- 264

35 Query: 297 QKAPTATPVAQPASTTNAVAAHPEAGLQPHVAAYKEKVASTYGVNEFSTYRAGDPGDHG 356
 AP A +P NAGLQP AA+KE+VAS +G+ FS YR GDPGDHG
 Sbjct: 265 SYAPNH-----AYNPMNAGLQPQTAAFKEEVASAFCITSFSGYRPGDPGDHG 311

40 Query: 357 KGLAVDFIVGTNQALGNKVAQYSTQNMAANNISYVIWQQKFYSNTNSIYGPANTWNAMPD 416
 KGLA+DF+V N ALG++VAQY+ +MA ISYVIW+Q+FY+ SIYGPA TWN MPD
 Sbjct: 312 KGLAIDFMVPENSALGDQVAQYAIDHMAERGILSYVIWKQRFYAPFASIYGPAYTWNMPD 371

45 Query: 417 RGGVTANHYDHVHVSFN 433
 RG +T NYDHVHVSFN
 Sbjct: 372 RGSITENHYDHVHVSFN 388

A related GBS gene <SEQ ID 8539> and protein <SEQ ID 8540> were also identified. Analysis of this protein sequence reveals the following:

45 Lipop: Possible site: -1 Crend: 3

 SRCFLG: 0

 McG: Length of UR: 20
 Peak Value of UR: 1.96
 Net Charge of CR: 2

50 McG: Discrim Score: 2.95
 GvH: Signal Score (-7.5): 3.84
 Possible site: 23
 >>> Seems to have a cleavable N-term signal seq.

55 Amino Acid Composition: calculated from 24
 ALOM program count: 0 value: 4.29 threshold: 0.0
 PERIPHERAL Likelihood = 4.29 58

modified ALOM score: -1.36

60 *** Reasoning Step: 3

Rule gpol

----- Final Results -----

65 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

SEQ ID 8540 (GBS322) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 77 (lane 9; MW 52kDa). The GBS322-His fusion product was purified (Figure 5 214, lane 10) and used to immunise mice. The resulting antiserum was used for FACS (Figure 267), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 296

10 A DNA sequence (GBSx0324) was identified in *S.agalactiae* <SEQ ID 955> which encodes the amino acid sequence <SEQ ID 956>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
    INTEGRAL      Likelihood = -1.86      Transmembrane      5 - 21 ( 4 - 21)
15
----- Final Results -----
bacterial membrane --- Certainty=0.1744 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

20 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC46072 GB:U50357 zoocin A endopeptidase [Streptococcus
zooepidemicus]
Identities = 163/274 (59%), Positives = 196/274 (71%), Gaps = 11/274 (4%)
25
Query: 25 VLADTYVRPIDNNGRITTGFNGYPGHCVDYAVPTGTIIIRAVADGTVKFAGAGANFSWMTD 84
V A TY RP+D G ITTGFNGYPGH GVDYAVP GT +RAVA+GTVKFA G AN WM
Sbjct: 21 VSAATYTRPLDTGNITTGFNGYPGHVGVDYAVPVGTPVRAVANGTVKFAGNGANHPWMLW 80
30
Query: 85 LAGNCVMIQHADGMHSGYAHMSRVARTGEVKQGDIIIGYVGATGMATGPHLHFELPAN 144
+AGNCV+I QHADGMH+GYAH+S++ T VKQG IIIG Y GATG TGPHLHF E LPAN
Sbjct: 81 MAGNCVLIQHADGMHTGYAHLSKISVSTDSTVKQGQIIGYTGATGQVTGPHLHFEMLPAN 140
35
Query: 145 PNFQNGFHGRINPTS LIANVATFSGKTQASAPS IKPLQSAPVQNQSSKLKVYRVDELQKV 204
PN+QNGF GRI+PT IAN F+G T + P N LK+Y+VD+LQK+
Sbjct: 141 PNWQNGFSGRIDPTGYIANAPVFNGTTPTE-----PTPTTTN---LKIYKVDDLQKI 189
Query: 205 NGVWLVKNNNTLTPTGFDWNDNGIPASEIDEVDANGNLTAQVLQKGGYFIFNPKTLKTVE 264
NG+W V+NN L PT F W DNGI A ++ EV +NG T+DQVLQKGGYF+ NP +K+V
40
Sbjct: 190 NGIWQVRNNNILVPTDFTWVDNGIAADDVIEVTSNGTRTSQVLQKGGYFVINPNNVKSVG 249
Query: 265 KPIQGTAGLTWAKTRFANGSSVWLRVDNSQELLY 298
P++G+ GL+WA+ F G +VWL + LLY
Sbjct: 250 TPMKGSGGLSWAQVNFTTGGNVWLNTTSKDNLLY 283
```

45 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8541> and protein <SEQ ID 8542> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1 Crend: 6
50
McG: Discrim Score: 6.63
GvH: Signal Score (-7.5): -2.97
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
ALOM program count: 1 value: -1.86 threshold: 0.0
55
    INTEGRAL      Likelihood = -1.86      Transmembrane      5 - 21 ( 4 - 21)
    PERIPHERAL    Likelihood = 5.57      50
```

modified ALOM score: 0.87

*** Reasoning Step: 3

5 ----- Final Results -----

bacterial membrane --- Certainty=0.1744 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

GP|2804351|gb|AAC46072.1||U50357(21 - 283 of 285) zoocin A endopeptidase {Streptococcus zoopidemicus}

%Match = 34.2

%Identity = 61.3 %Similarity = 74.4

15 Matches = 163 Mismatches = 65 Conservative Sub.s = 35

144	174	204	234	264	294	324	354
VV*VFLS*LRYTTYILKTFLIKPKYSSR*	VLF	LIF*	FKFSNKLIA	SV*ALHYIN	SIWRFFLN	KWLVKASSL	VVLGGMV

20 MKRIFFAFLSLCLF
10

384	414	444	474	504	534	564	594
LSAGSRVLA	D	T	V	R	P	I	D
25 T	A	D	T	V	R	P	I
IFGTQTV	S	A	A	T	R	P	I
30 G	T	R	P	L	D	T	G
HSGYA	H	M	S	R	V	V	A
35 T	S	R	V	V	A	T	G
HTGYAH	L	S	K	I	S	V	T
40 K	I	S	V	S	D	S	T
110	120	130	140	150	160		

624	654	684	714	744	774	804	834
HSGYAH	M	S	R	V	V	A	G
30 G	S	R	V	V	A	T	A
HTGYAH	L	S	K	I	G	T	G
35 T	I	S	V	S	G	T	N
110	120	130	140	150	160		

864	894	924	954	984	1014	1044	1074
KPLQSAPV	Q	NQSSKL	KVYRVDELQKVNGVWLVKVNNTLTPTGF	WDNDNGIPASEIDEVDANGNL	TADQVLQKG	GGYFIFNPK	
40 K	Q	S	K	L	N	G	
180	190	200	210	220	230	240	

1104	1134	1164	1194	1224	1254	1284	1314	
TLKTVEKPIQGTAGLTWAKTRF	A	GSSVWL	RVDNSQELLYK*	FEVLIHC	FK*	QLCY*	LSTISLNRLKII	L*SSKV*YYSL
45 G	P	M	K	Q	N	I	I	
260	270	280						

SEQ ID 8542 (GBS36) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 4; MW 34.1kDa).

GBS36-His was purified as shown in Figure 192, lane 7.

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 297

55 A DNA sequence (GBSx0325) was identified in *S.agalactiae* <SEQ ID 957> which encodes the amino acid sequence <SEQ ID 958>. This protein is predicted to be phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohyd. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have no N-terminal signal sequence

-380-

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2815 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5

The protein has homology with the following sequences in the GENPEPT database:

>GP: BAB04352 GB: AP001509 phosphoribosylaminoimidazolecarboxamide
 formyltransferase/IMP cyclohydrolase [Bacillus halodurans]
 Identities = 310/515 (60%), Positives = 390/515 (75%), Gaps = 4/515 (0%)

10

Query: 1 MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDDVTGFPEMMD 60
 M +RAL+SVS+K GI+ FAK L +I+STGGTK AL +AG+ I DVTGFPE++D
 Sbjct: 1 MKRRAIVSVSNKEGIVPFAKALVEHEVEI VSTGGTKRALQEAGIPVIGISDVTFPEILD 60

15

Query: 61 GRVKTLHPNIHGGLLARRDADSHLQAACKDNIELIDL VVVVNLYPFKETILRPDV TYDLAV 120
 GRVKTLHPNIHGGLLA R+ D HL +++I ID VVVNLYPF++TI +P+ T+ A+
 Sbjct: 61 GRVKTLHPNIHGGLLAMRERDEHLAQLINEHHIRPIDFVVVNLYPFQQTIAKPEATFADAI 120

20

Query: 121 ENIDIGGPSMLRSAAKNHASVTVVVDSADYATVVLGELADASQTTFKTRQLAAKA FRHTA 180
 ENIDIGGPSMLR+AAKNH VTVVVD DY TVL ELAD +T++RLAAK FRHTA
 Sbjct: 121 ENIDIGGPSMLRAAKNHQHVTVVVVDVYETV LKELADQGNVATEKRLAAKVFRHTA 180

25

Query: 181 AYDALIAEYFTAQVGEAKPEKLTITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240
 AYDA+IAEY T VGE PE LT+T++ KQ +RYGENP Q A FYQK L SIA AKQ
 Sbjct: 181 AYDAMIAEYLTDAGVEESPESLTVTPEKKQDLRYGENPQHOKATFYQKPLGAKASTAHAKQ 240

30

Query: 241 LNGKELSFNNIRDADAIRIIRDFKDSPTVVALKHMNP CGIGQADDIETAWDYAYEADPV 300
 L+GKELS+NNI DADAA+ I+++FK+ P VA+KHMNP CG+G + I+ A+D AYEADPV
 Sbjct: 241 LHGKELSYNNINDADAALSIVKEFKE-PAAVAVKHMNP CGVG TGETIKEAFDKAYEADPV 299

35

Query: 301 SIFGGIVVLNREVDAATAEKMHPIFL IIIAPS YSEEALAILTNKKKNLRILELPFDAQA 360
 SIFGGI+ LNREVD TA+ + IFLE IIIAPS+SSEEAL +LT+ KKNLR+L LP + +
 Sbjct: 300 SIFGGII ALNREVDTV EAKTLKEIFL IIIAPS FSEEAL DV LTS-KKNLRLLTLPLNEE- 357

40

Query: 361 ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG 420
 ++ E T + GG LVQ +D ++ ++ T R+PTE E AL+ AW+ +K+VKS N
 Sbjct: 358 -NQAEKRITSIHGGALVQEEEDTYGFEEAEIKIPTKREPTAEWEALKLAWRVVKHVKSNA 416

Query: 421 IIIITNDHMTLGLGAGQTNR VGSVKIAIEQAKDHLDGAVLASDAFFPFDADNIEEIAAGIK 480
 I++ + MT+G+GAGQ NR VG+ KIAIEQ A + G+V+ SDAFFP D +E A AGI
 Sbjct: 417 IVLADGQM TGVGAGQ MN RVGA KIAIEQAGEKAAGSVMGSDAFFPMGDTVELAAKAGIT 476

45

Query: 481 AIIQPGGSVRDQESIDAANKHGLTMIFTGVRHF RH 515
 AIIQPGGS+RD+ESI+ A+KHG+ M+FTGVRHF+H
 Sbjct: 477 AIIQPGGSIRDEESIENADKHGTIAMVFTGVRHF KH 511

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 959> which encodes the amino acid sequence <SEQ ID 960>. Analysis of this protein sequence reveals the following:

Possible site: 48

50

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2932 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55

An alignment of the GAS and GBS proteins is shown below:

Identities = 500/515 (97%), Positives = 507/515 (98%)

60

Query: 1 MTKRALISVSDKSGIIDFAKELKNLGWDIISTGGTKVALDDAGVETIAIDDVTGFPEMMD 60
 MTKRALISVSDKSGI+DFAKELKNLGWDIISTGGTKV LDDAGVETIAIDDVT FPEMMD
 Sbjct: 1 MTKRALISVSDKSGIVDFAKELKNLGWDIISTGGTKVTLDDAGVETIAIDDVTGFPEMMD 60

Query: 61 GRVKTLHPNIHGGLLARRDADSHLQAACKDNIELIDL VVVVNLYPFKETILRPDV TYDLAV 120

-381-

GRVKTLHPNIHGGLLARRDADSHLQAAKDNIELIDLVVVNLYPFKETILRPD+TYDLAV
 Sbjct: 61 GRVKTLHPNIHGGLLARRDADSHLQAAKDNIELIDLVVVNLYPFKETILRPDITYDLAV 120

Query: 121 ENIDIGGPMRLSAAKNHSVTVVVDSADYATVLGELADASQTTFKTRQRLLAAKAFRHTA 180
 ENIDIGGPMRLSAAKNHSVTVVV ADYATVLGELADA QTTF+TRQRLLAAK FRHTA
 Sbjct: 121 ENIDIGGPMRLSAAKNHSVTVVVPADYATVLGELADAGQTTFETRQRLLAAKVFRHTA 180

Query: 181 AYDALIAEYFTAQVGEAKPEKLITITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240
 AYDALIAEYFT QVGEAKPEKLITITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ
 Sbjct: 181 AYDALIAEYFTTQVGEAKPEKLITITYDLKQAMRYGENPQQDADFYQKALPTDYSIASAKQ 240

Query: 241 LNGKELSFnNIRDAAAIRIIRDFKDSPTVVALKHMNPCCIGQADDIETAWDY Y+ADPV 300
 LNGKELSFnNIRDAAAIRIIRDFKD PTVALKHMNPCCIGQADDIETAWDY Y+ADPV
 Sbjct: 241 LNGKELSFnNIRDAAAIRIIRDFKDPTVALKHMNPCCIGQADDIETAWDY Y+ADPV 300

Query: 301 SIFGGIVVILNREVDAATAEKMHPIFLEIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360
 SIFGGI+VNLREVDAATA+KMHPIFLEIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360
 Sbjct: 301 SIFGGIIVILNREVDAATAKKMHPIFLEIIIAPSYSEEALAILTNKKKNLRILELPFDAQA 360

Query: 361 ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG 420
 ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG
 Sbjct: 361 ASEVEAEYTGVVGGLLVQNQDVVAENPSDWQVVTDRQPTEQEATALEFAWKAIKYVKSNG 420

Query: 421 IIIITNDHMTLGLGAGQTNRGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAGIK 480
 IIIITNDHMTLGLGAGQTNRGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAGIK
 Sbjct: 421 IIIITNDHMTLGLGAGQTNRGSVKIAIEQAKDHLDGAVLASDAFFPFADNIEEIAAGIK 480

Query: 481 AIIQPQGGSVRDQESIDAANKHGLTMIFTGVRHFRH 515
 AIIQPQGGSVRDQ+SIDAANKHGLTMIFTGVRHFRH
 Sbjct: 481 AIIQPQGGSVRDQDSIDAANKHGLTMIFTGVRHFRH 515

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 298

35 A DNA sequence (GBSx0326) was identified in *S.agalactiae* <SEQ ID 961> which encodes the amino acid sequence <SEQ ID 962>. This protein is predicted to be similar to antibiotic resistance protein. Analysis of this protein sequence reveals the following:

Possible site: 46
 >>> Seems to have no N-terminal signal sequence
 40 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1842 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12342 GB:Z99106 similar to antibiotic resistance protein
 [Bacillus subtilis]
 Identities = 65/263 (24%), Positives = 117/263 (43%), Gaps = 34/263 (12%)
 50 Query: 5 KNLEIVESIFGD-WDETIWSCV-QGIMGEVFVDSLDPKSSLAKLGRKSSFGFLAGQPT 62
 K +--+F D + T ++S + Q I G V+ D PKS +G +S F+AG
 Sbjct: 10 KKYSSLKTMFDKCYCPTFVYSILDQTFPKSFF--IGTESGIYFIAGDQG 67

55 Query: 63 -----LFLLEVCSGEDIILVQHKGWSDLIESTYGQNAHSFKRYATKKDTLFERS 112
 + +V S + L W +++ + + R A +
 Sbjct: 68 NRDFHDFIAGYYEEQVKSSKRFTLFSSSDTWDSVLKPILKDDLNMQRRAAFSY----QP 122

60 Query: 113 RLEKFVTLQLPNPGFELRAIDEKV-----YNSCLEKEWSQDLVANYATYQYYKKQGIGYVV 166
 + K QLP G L+ IDE + +NS +E+ + + + +G G+ V
 Sbjct: 123 KSFKKTLQLPKGLVLKRIDEIISHSTAFNSAYYEY-----WNSVSQFASKGFGFAV 175

Query: 167 YYQGMIIAGASSYSTYKNGIEIEVDTHPDFRRGLATIVAAQLILTCLDKGIYPSWDAH- 225
 + +++ +S N E++ T ++R GLA VA + I C++ GI PSWD

Sbjct: 176 LHGNHVVSECTSIFLGHNRAEMDIYTLEEYRGLGLAYCVANRFIAFCMENGIVPSWDCDI 235

5

Query: 226 -TRTSNLSEKLGYESHEYIAY 247
 +S+ L+ KLG++ EY Y

Sbjct: 236 CNNSSIALAKLGFKTVTEYTIY 258

10 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 299

15 A DNA sequence (GBSx0328) was identified in *S.agalactiae* <SEQ ID 963> which encodes the amino acid sequence <SEQ ID 964>. This protein is predicted to be phosphoribosylglycinamide formyltransferase homolog (purN). Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.0736 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 965> which encodes the amino acid sequence <SEQ ID 966>. Analysis of this protein sequence reveals the following:

Possible site: 48
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.53 Transmembrane 75 - 91 (75 - 91)

30 ----- Final Results -----

bacterial membrane --- Certainty=0.1213 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has homology with the following sequences in the databases:

>GP:CAA04374 GB:AJ000883 purD [Lactococcus lactis]
 Identities = 236/419 (56%), Positives = 301/419 (71%), Gaps = 7/419 (1%)

40 Query: 50 LKLLVVGGREHAIAKKLLASKGVDFQVFVAPGNDGMLDGLDLVNIVVSEHSRLIAFAK 109
 +K+LV+GSGGREHA+AKK + S V++VFVAPGN GM DG+ +V+I + +L+ FA+
 Sbjct: 1 MKILVIGSGGREHALAKKFMEQPVEEVFVAPGNSGMEKDGIQIVHISELSNDKLVKFAQ 60

45 Query: 110 ENEISWAFIGPDDALAAGIVDDFNSAGLRAFGPTKAAAEELEWSKDFAKEIMVKYNVPTAA 169
 I F+GP+ AL G+VD F A L FGP K AAELE SKDFAK IM KY VPTA
 Sbjct: 61 NQNIGLTFVGVPETALMNGVVDAFIKAELPIFGPNKMAAELEGSKDFAKSIMKKYGVPTAD 120

50 Query: 170 YGTFSDFEKAKAYIEEQGAPIVVKADGLALGKGVVVAETVEQAVEAAQEMLLDNKFGDSG 229
 Y TF E A AY++E+G P++V+KADGLA GKGV VA +E A A ++ F S
 Sbjct: 121 YATFDSLEPALAYLDEKGVPVLVIKADGLAAGKGVTAFDIETAKSALADI----FSGSQ 175

55 Query: 230 ARVVIIEFLDGEESLFAFANGDKFYIMPTAQDHKRAFDGDKGPNNTGGMGAYAPVPHLPQ 289
 +VVIEEFLDGEESLSF+F + K Y MP AQDHKRAFD DKGPNTGGMGAY+PV H+ +
 Sbjct: 176 GKVVIEEFLDGEESLFSFIHDGKIYPMPIAQDHKRAFDKDGPNTGGMGAYSPVLHISK 235

Query: 290 SVVDTAVEMIVRPVLEGMVAEGRPYLGVLYVGLILITADGPKVIEFNSRGDPETQIILPR 349
 VV+ A+E +V+B + GM+ EG+ GVLY GLILT DG K IEFN+RFGDPETQ++LPR
 Sbjct: 236 EVVNEALEKVVVKPTVAGMIEEGKSFTGVLYAGLILTEDGVKTIEFNARFGDPETQVVLPR 295

Query: 350 LTSDFAQNIDDDIMMGIEPYITWQKDGVTLGVVVASEGYPFDYEKGVPLPEKTDGDIITYYY 409
 L SD AQ I DI+ G EP + W + GVTLGVVVA+EGYP + G+ LPE +G + YY

Sbjct: 296 LKSDLAQAIIDILAGNEPLEWESGVTLGVVVAAEGYPSQAKLGLLPEIPEG-LNVYY 354

5

Query: 410 AGVKFSENSELLLSNGGRVYMLVTTEDSVKAGQDKIYTQLAQQDTTGLFYRNDIGSKAI 468
 AGV +EN++ L+S+GGRVY++ T + VK+ Q +Y +L + + G FYR+DIGS+AI

Sbjct: 355 AGVSKNENNQ-LISSGGRVYLVSETGEDVKSTQKLLYEKLDKLENDGFYRHDIGSRAI 412

10 An alignment of the GAS and GBS proteins is shown below:

Identities = 172/182 (94%), Positives = 176/182 (96%)

Query: 1 MKIAVFASGNQNSNFQVIAEQFQVSFVFSDHARDAYVLERAQNLAI P S F A F E L K E F E N K A A Y 60
 MKIAVFASGNQNSNFQVIAEQF VSFVFSDHARDAYVLERAQNLAI P S F A F E L K E F E N K A Y

15

Sbjct: 1 MKIAVFASGNQNSNFQVIAEQFPVSFVFSDHARDAYVLERAQNLAI P S F A F E L K E F E N K V A Y 60

Query: 61 EQAVVDLLDKHEIDLVCCLAGYMKIVGETLLSAYEGRIINIHPTYLPEFGAHLGIKDAWEA 120
 EQA+VDLLDKHEIDLVCCLAGYMKIVGETLL AYE RIINIHP YLPEFGAHLG+DAWEA

20

Sbjct: 61 EQAIVDLDKHEIDLVCCLAGYMKIVGETLLAYERRIINIHPAYLPEFGAHLGIEDAWEA 120

Query: 121 GVDQSGVTIHWVDSGVDTGQVIQQVHVPRLADDLESFETRIHETEYQLYPAVLDSLGIK 180
 GVDQSGVTIHWVDSGVDTGQVIQQV VPRLADDLESFETRIHETEYQLYPAVLDSLGIK++

Sbjct: 121 GVDQSGVTIHWVDSGVDTGQVIQQVRVPRLADDLESFETRIHETEYQLYPAVLDSLGIK 180

25

Query: 181 RK 182

RK

Sbjct: 181 RK 182

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for 30 vaccines or diagnostics.

Example 300

A DNA sequence (GBSx0329) was identified in *S.agalactiae* <SEQ ID 967> which encodes the amino acid sequence <SEQ ID 968>. Analysis of this protein sequence reveals the following:

Possible site: 52

35

>>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.59 Transmembrane 121 - 137 (121 - 137)

----- Final Results -----

40

bacterial membrane --- Certainty=0.1235 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP: AAC16901 GB: AF016634 phosphoribosylformylglycinamide

45

cyclo-ligase [Lactococcus lactis subsp. cremoris]

Identities = 253/338 (74%), Positives = 288/338 (84%), Gaps = 4/338 (1%)

Query: 4 KNAYAQSGVDVEAGYEVVERIKKHVARTERAGVMALGGFGGMFDLSQTGVKEPVLIISGT 63

+NAYA+SGVDEAGYEVV RIKKHVA+TER GV+GALGGFGG FDLS VKEPVLIISGT

50

Sbjct: 5 ENAYAKSGVDVEAGYEVVSRIKKHVAKTERLGVLGALGGFGGSFDLSVLDVKEPVLIISGT 64

Query: 64 DGVGTKLMLAIIKYDKHTIGQDCVAMCVNDIIAAGAEPLYFLDYVATGKNEPAKLEQVVA 123

DGVGTKLMLAI+ DKHTIG DCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLEQVVA

55

Sbjct: 65 DGVGTKLMLAIIRADKHTIGIDCVAMCVNDIIAAGAEPLYFLDYIATGKNIPEKLEQVVA 124

Query: 124 GVAEGCVQASAALIGGETAEMPGMYGEDDYDLAGFAVGVAEKSQIIDGSK-VKEGDILLG 182

GVAEGC+QA AALIGGETAEMPGMY EDDYDLAGFAVGVAEKSQ+IDG K V+ GD+LLG

Sbjct: 125 GVAEGCLQAGAAALIGGETAEMPGMYDEDDYDLAGFAVGVAEKSQIIDGKDVEAGDVLLG 184

60

Query: 183 LASSGIHSNGYSLVRRVFADYTGDEVLP ELEGKQLKDVLLEPTRIYVKAALPLIKEELVN 242

LASSGIHSNGYSLVR+VFAD+ +E LPEL+ + L D LL PT+IYVK LPLIK+ +

-384-

Sbjct: 185 LASSGIHSNGYSLVRKVFADFDLNESLPELD-QSLIDTLLPTKIVVKELLPLIKQNLIK 243
 Query: 243 GIAHITGGGFIENVPRMFADDLAAEIDEDKVPVLPPIFKALEKYGDIKHEEMFEIFNMGVG 302
 GIAHITGGGF EN+PRMF + L+AEI E VLPPIFKALEKYG IKHEEM+EIFNMG+G
 Sbjct: 244 GIAHITGGGFHENLPRMFGNSLSAEIVEGSWDVLPPIFKALEKYGSIKHEEMYEIFNMGIG 303
 Query: 303 LMLDVNPENVDRVKELLDEPVYEIGRIIKKADDSVVIK 340
 +++ V PEN +K+ L+ +EIG+++ + + VVIK
 Sbjct: 304 MVIAVAPENAAALKELN--AFEIGQMVNRQEAPVVIK 339

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 969> which encodes the amino acid sequence <SEQ ID 970>. Analysis of this protein sequence reveals the following:

Possible site: 38
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3236 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20 An alignment of the GAS and GBS proteins is shown below:

Identities = 321/340 (94%), Positives = 332/340 (97%)

25 Query: 1 MSEKNAYAQSGVDVEAGYEVVERIKHVARTERAGVMGALGGFGGMFDLSQTGVKEPVLI 60
 MSEKNAYA+SGVDVEAGYEVVERIKHVARTERAGVMGALGGFGGMFDLS+TGVKEPVLI+
 Sbjct: 1 MSEKNAYAKSGVDVEAGYEVVERIKHVARTERAGVMGALGGFGGMFDLSKTGVKEPVLI 60
 Query: 61 SGTDGVGTKLMLAIKYDKHDТИQDCVAMCVNDIIAAGAEPLYFLDYVATGKNEPAKLEQ 120
 SGTDGVGTKLMLAIKYDKHDТИQDCVAMCVNDIIAAGAEPLYFLDY+ATGKN P KLE+
 Sbjct: 61 SGTDGVGTKLMLAIKYDKHDТИQDCVAMCVNDIIAAGAEPLYFLDYIATGKNNPVKLEE 120
 Query: 121 VVAGVAEGCVQASAALIGGETAEMPGMYGEDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180
 VV+GVAEGCVQA AALIGGETAEMPGMYG+DDYDLAGFAVGVAEKSQIIDGSKVKEGDIL
 Sbjct: 121 VVSGVAEGCVQAGAACALIGGETAEMPGMYGQDDYDLAGFAVGVAEKSQIIDGSKVKEGDIL 180
 35 Query: 181 LGlassgihsngySLVRRVFADYTGDEVLPLELEGKQLKDVLLEPTRIYVKAALPLIKEEL 240
 LGlassgihsngySLVRRVFADYTG E+LPLELEGKQLKDVLLEPTRIYVKAALPLIKEEL
 Sbjct: 181 LGlassgihsngySLVRRVFADYTGKELLPELEGKQLKDVLLEPTRIYVKAALPLIKEEL 240
 40 Query: 241 VNGIAHITGGGFIENVPRMFADDLAAEIDEDKVPVLPPIFKALEKYGDIKHEEMFEIFNMG 300
 V GI HITGGGFIEN+PRMFADDLAAEIDEDKVPVLPPIFKALEKYGDIKHEEMFEIFNMG
 Sbjct: 241 VKGIGHITGGGFIENIPRMFADDLAAEIDEDKVPVLPPIFKALEKYGDIKHEEMFEIFNMG 300
 45 Query: 301 VGLMLDVPENVDRVKELLDEPVYEIGRIIKKADDSVVIK 340
 VGLML V+PENV+RVKELLDEPVYEIGRIIKKAD SVVIK
 Sbjct: 301 VGLMLAVSPENVRVKEELLDEPVYEIGRIIKKADASVVIK 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 301

A DNA sequence (GBSx0330) was identified in *S.agalactiae* <SEQ ID 971> which encodes the amino acid sequence <SEQ ID 972>. This protein is predicted to be phosphoribosylpyrophosphate amidotransferase (purF). Analysis of this protein sequence reveals the following:

Possible site: 25
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1112 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

-385-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

5 >GP:AAD12627 GB:U64311 phosphoribosylpyrophosphate amidotransferase
 [Lactococcus lactis]
 Identities = 340/470 (72%), Positives = 404/470 (85%), Gaps = 6/470 (1%)

 10 Query: 3 YEVKSLNEECVFGIWGYPQAAQVTYFGLHSLQHRGQEGAGIISNDNGKLYGYRNVGLLS 62
 +E K+LNEECG+FG+WG+P AA++TYFGLH+LQHRGQEGAGI+ N+NGKL +R +GL++
 Sbjct: 37 FEAKTLNEECGLFGVWGHPDARLTYFGLHALQHRGQEGAGILVNNNGKLNRRHGLGLVT 96

 15 Query: 63 EVFKNQSELDNLTGNAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHGNLTNAISS 122
 EVF+++ L+ LTG++AIGHVRYATAGSA+I NIQPF ++FHDG L HGNLTNA S
 Sbjct: 97 EVFRIEKDLLELTGSSAIGHVRYATAGSANINNIQPQFQEFHDGSLGLAHNGNLTNAQSL 156

 20 Query: 123 RKELEKQGAIFNASSDTEILMHLIRRSHNPNSFMGKVKEALSTVKGGFAYLLMTEDKLIAA 182
 R ELEK GAIF+++SDTEILMHLIRRSH+P FMG+VKEAL+TVKGGFAYL+MTE+ ++AA
 Sbjct: 157 RCELEKSGAIFSSNSDTEILMHLIRRSHHPFMGRVKEALNTVKGGFAYLIMTENSIVAA 216

 25 Query: 183 LDPNAFRPLSIGQMONGAWVISSETCAFEVVGAKWVVDVEPGEVILIDDSGIQCDRYTDE 242
 LDPN FRPLSIG+M NGA V++SETCAF+VVGA W++DV+PGE+I I+D GI D++TD
 Sbjct: 217 LDPNNGFRPLSIGKMSNGALVVASETCAFDVVGATWIQDVQPGIEIIIEINDDGIHVDQFTDS 276

 30 Query: 243 TQLAICSMYEVVYFARPDSIHTGVNVHTARKNMGKRLAQEFKDADIVIGVPNSSLSAAMG 302
 T + ICSMEY+YFARPDS I GVNVHTARK GK LAQE K DADIVIGVPNSSLSAA G
 Sbjct: 277 TNMTICSMYEIYFARPDSNIAGVNVHTARKRSKGKILAQEAKIDADIVIGVPNSSLSAASG 336

 35 Query: 303 FAEESGLPNEMLVKNQYTORTFIQPTQELREQGVRMKLSAVSGVVKGKRVVMIDDSIVR 362
 +AEEESGLP EMGL+KNQY RTFIQPTQELREQGVRMKLSAV GVV+GKRV+M+DDSVIR
 Sbjct: 337 YAEEESGLPYEMGLIKNQYVARTFIQPTQELREQGVRMKLSAVRGVVEGKRVIMVDDSVIR 396

 40 Query: 363 GTTSRRIVGLLREAGATEVHVIAIASPELKYPFCFYGIDIQTRRELISANHAVDEVCDIIGA 422
 GTTSRRIV LL++AGA EVHVAIASP LKYPFCFYGIDIQ R ELI+A H DE+ + IGA
 Sbjct: 397 GTTSRRIVKLLKDAGAAEVHVIAIASPALKYPFCFYGIDIQDRDELIAATHITDEIREAIGA 456

 45 Query: 423 DSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPTPLYDVEEYLRSL 472
 DSLTYLS GL+++IG + LC++YFDG YPTPLYDYE +YL SL
 Sbjct: 457 DSLTYLSQSGLVIAIG-----HDKLCLSYFDGEYPTPLYDYEADYLESL 500

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 973> which encodes the amino acid sequence <SEQ ID 974>. Analysis of this protein sequence reveals the following:

Possible site: 21
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.0610 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 An alignment of the GAS and GBS proteins is shown below:

Identities = 473/484 (97%), Positives = 481/484 (98%)

Query: 1 MTYEVKSLNEECVFGIWGYPQAAQVTYFGLHSLQHRGQEGAGIISNDNGKLYGYRNVGL 60
 MTYEVKSLNEECVFGIWG+PQAAQVTYFGLHSLQHRGQEGAGI+SNDNGKLYGYRNVGL
 Sbjct: 20 MTYEVKSLNEECVFGIWGPQAAQVTYFGLHSLQHRGQEGAGIVSNDNGKLYGYRNVGL 79

 Query: 61 LSEVFKNQSELDNLTGNAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHGNLTNAI 120
 LSEVFKNQSELDNLTGNAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHGNLTNAI
 Sbjct: 80 LSEVFKNQSELDNLTGNAIGHVRYATAGSADIRNIQPFLYKFHDGQFALCHGNLTNAI 139

 60 Query: 121 SSRKELEKQGAIFNASSDTEILMHLIRRSHNPNSFMGKVKEALSTVKGGFAYLLMTEDKLI 180
 S RKELEKQGAIFNASSDTEILMHLIRRSHN SFMGKVKEAL+TVKGGFAYLLMTE+KLI
 Sbjct: 140 SLRKELEKQGAIFNASSDTEILMHLIRRSHNSSFMSGKVKEALNTVKGGFAYLLMTEKLI 199

-386-

Query: 181 AALDPNAFRPLSIGMQNGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQCDRYT 240
 AALDPNAFRPLSIGMQNGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDSGIQCDRYT
 Sbjct: 200 AALDPNAFRPLSIGMQNGAWVISSETCAFEVVGAKWVRDVEPGEVILIDDRGIQCDRYT 259

5 Query: 241 DETQLAICSMEYVYFARPDSIHGvnVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSSAA 300
 DETQLAICSMEYVYFARPDSIHGvnVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSSAA
 Sbjct: 260 DETQLAICSMEYVYFARPDSIHGvnVHTARKNMGKRLAQEFKQDADIVIGVPNSSLSSAA 319

10 Query: 301 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMQLSAVSGVVKGRVVMMDDSI 360
 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMQLSAVSGVVKGRVVMMDDSI
 Sbjct: 320 MGFAEESGLPNEMGLVKNQYTQRTFIQPTQELREQGVRMQLSAVSGVVKGRVVMMDDSI 379

15 Query: 361 VRGTTSRIVGLLREAGATEVHVAIASPELKYPFCYGIDIQTRRELISANHAVDEVCDII 420
 VRGTTSRIVGLLREAGA+EVHVAIASPELKYPFCYGIDIQTRRELISANH+VDEVCDII
 Sbjct: 380 VRGTTSRIVGLLREAGASEVHVAIASPELKYPFCYGIDIQTRRELISANHSVDEVCDII 439

20 Query: 421 GADSLTYLSIDGLIKSIGLETKAPNGGLCVAYFDGHYPPTLYDYEEEYLRSLEEKTSFYI 480
 GADSLTYLS+DGLI+SIGLETKAPNGGLCVAYFDGHYPPTLYDYEEEYLRSLEEKTSFYI
 Sbjct: 440 GADSLTYLSLDGLIESIGLETKAPNGGLCVAYFDGHYPPTLYDYEEEYLRSLEEKTSFYI 499

Query: 481 QKVK 484
 QKVK
 Sbjct: 500 QKVK 503

25 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 302

A DNA sequence (GBSx0331) was identified in *S.agalactiae* <SEQ ID 975> which encodes the amino acid sequence <SEQ ID 976>. Analysis of this protein sequence reveals the following:

30 Possible site: 28
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4797 (Affirmative) < succ>
 35 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

40 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 303

A DNA sequence (GBSx0332) was identified in *S.agalactiae* <SEQ ID 977> which encodes the amino acid sequence <SEQ ID 978>. Analysis of this protein sequence reveals the following:

45 Possible site: 13
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3489 (Affirmative) < succ>
 50 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 304

A DNA sequence (GBSx0333) was identified in *S.agalactiae* <SEQ ID 979> which encodes the amino acid sequence <SEQ ID 980>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

10 bacterial cytoplasm --- Certainty=0.1690(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP: CAC12194 GB: AL445066 phosphoribosylformylglycinamide synthase
 related protein [Thermoplasma acidophilum]
 Identities = 199/746 (26%), Positives = 329/746 (43%), Gaps = 103/746 (13%)
 Query: 202 ADD--FAAYKAEQGLAMEVDDLLFIQDYFKSIGRVPTETELKVLDTYWSDHCRHTTFETE 259
 ADD A GLA+ D++ ++ YF+ +GR P + E+ + WS+HC + + +
 Sbjct: 11 ADDARLKAISKRGLALSLDEMKA VRSYFERLGRDPIDAEIHAV AQSWEHCSYKSSKYY 70
 Query: 260 LKNIDFSASKFQKQLQATYDKYIAMRDELGRSEKPQTLM DMATIFGRYERANGRLDDMEV 319
 LK K+ L+ Y +AM D+ G
 Sbjct: 71 LK-----KYL GSLKTDYT-ILAMEDDAG----- 92
 Query: 320 SDEINACSVEIEVDV DGKEPWLLMFKNETHNHPTEIEPFGGAATCIGGAI RDPLSGRSY 379
 VD DG + + K E+HNHP+ +EP+GGAAT IGG +RD L +
 Sbjct: 93 -----VVDFDG---EYAYVLKMESHNHPSAVEPYGGAATGIGGI VRDVLCMGAQ 138
 Query: 380 VYQAMRISGAGDDITTPIAETRAGKLPQQVISKTAAHGYSYGNQIGLATTYVREYFHPGF 439
 + GD+++ E G L + I G YGN+IG+ YF +
 Sbjct: 139 PVALIDSLFLG DVSSDRYE---GLLSPRYIFGGV VGGI RDYGNRIGIPNVAGSLYFDKLY 195
 Query: 440 VAKRMELGAVVGAAPKENVVREKP-EAGDVV VLLGGKTGRDGV GGATGSSKVQTVESVET 498
 + + VG ++ +VR K + GDV+VL+GGKTGRDG+ G +S + + ++
 Sbjct: 196 NSNP LVNAGCVGIVRRDRIVRSKS YKPGDV LVMGGKTGRDGIHG VNFAS TT LG-KVTKS 254
 Query: 499 AGAEVQKGNAIEERKIQRLFRDGNVTR LIKKNSNDFGAGGV CVAIGELAD---GLEIDL D 554
 + + Q GN I E+ + + + N LI+ D G GG+ A E+ G EI LD
 Sbjct: 255 SRLAIQLGNPIVEQPMIKAVLEANDAGLIRAMKDLGGGLSSAATEMVYAGGFGAEITLD 314
 Query: 555 KVPLKYQGLNGTEIAISESQERM SVVVGPSDV AFIAACNKENIDAVVVATVTEKPNLVM 614
 + LK ++G EI ISESQERM + P DV+ K N+D V+ VT + +
 Sbjct: 315 DIKLKESNMSGWEIW ISESQERM LMECYPEDVEKIRQIAEKWNLDFSVIGQVTADRRIRV 374
 Query: 615 TWNGETIVDLERCFLDTNGV-RVVVDAKVKD LTVPEARTT SAETLEADM KVLS DLNH 673
 + I+D++ FLD + V + K V+K +TVP+ E L + + ++ LN
 Sbjct: 375 YYKKRKIIDMDIEFLDDSPVYQRPYRIKEVEKS VTV PQ----EPEDLNSFVRDFMARLNT 430
 Query: 674 ASQKGLQTIFDSSVGRSTVN HPIGGR-YQITPTESSVQ KLPVQYGVTTASVMAQGYN PY 732
 ++ + +D +V ST+ P GR + T ++ +V K P++ + V+ G P
 Sbjct: 431 CARFN VVRQYDHTVRGSTIVTPFVGRPNKETHADATVIK-PL ENSM--RGLVLTSGSRPN 487
 Query: 733 IAEWSPYHGAAV A VIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQP VSALL GSI 792
 + PY G + EA +++TG R ++ E GQ V ++
 Sbjct: 488 MVSVD P YAGTLLT LAEAYKNILSTG---GRPHS VVDALNFGNPEREEIMQ QF VESVRAIG 544
 Query: 793 EAQI QFGLPSI GGKDS MSGT FEEL TVPPTL VAFGV TTADS-RKVL SPEFKAAGENI Y--- 848
 + + GLP + G S + + + PT V D R+ + K +G IY
 Sbjct: 545 DF CRKMGLP VVAGNV SFYNEYRKT DIMPTPTIMMVGLIDD VRRSRTT YMKGSGN A IY LIG 604

Query: 849 -----YIPGQAISEDIDFDLIKANF--SQFEAIQAQHKITAASAVKYGG 890
 Y G + D+D +F S+ + I + H +++ GG
 Sbjct: 605 EPCDNLTCSEYSRMHGYTDGFLPAPDLDDELTRIRDFLSSKADMILSSHDVSS-----GG 658

5 Query: 891 VLESLALMTFGNRIGASVEIAELDSS 916
 + +L+ M+FG+ IG V+I+ + ++
 Sbjct: 659 LFAALSEMSFGSGIGFHVDISNVSA 684

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 981> which encodes the amino acid
 10 sequence <SEQ ID 982>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1415 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 1219/1256 (97%), Positives = 1226/1256 (97%)

Query: 11 SSYFRVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLASLKDLRIVQVYDVF 70

SSYF VAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQL SLK LRIVQVYDVF

Sbjct: 2 SSYFPVAPLSDLVSYMNKRIFVEKKADFGIKSASLVKELTHNLQLTSKLALRIVQVYDVF 61

25 Query: 71 NLAEDLLARAEEKHIFSEQVTDRLLTEAEITAEELDKVAFFAIEALPGQFDQRAASSQEALL 130

NLAEDLLARAEEKHIFSEQVTD LLTE ITAEELDKVAFFAIEALPGQFDQRAASSQEALL

Sbjct: 62 NLAEDLLARAEEKHIFSEQVTDCLLTETEITAEELDKVAFFAIEALPGQFDQRAASSQEALL 121

30 Query: 131 LLGSDSQVKVNNTAQQLYLVNKDI EAEAELEAVKVNLYLLNPVDSRFKDITLPLEVQAFSVSDKT 190

L GSDSQVKVNNTAQQLYLVNKDI EAEAELEAVKVNLYLLNPVDSRFKDITLPLE QAFSVSDKT

Sbjct: 122 LFGSDSQVKVNNTAQQLYLVNKDITEAEAELEAVKVNLYLLNPVDSRFKDITLPLEQAFSVSDKT 181

35 Query: 191 ISNLDFFETYQADDFAAYKAEGQGLAMEVDDLLFIQDYFKSIGRVPTETELKVLDTYWSDH 250

I NLDFFETYQADDFA YKAEGQGLAMEVDDLLFIQ+YFKSIG VPTETELKVLDTYWSDH

Sbjct: 182 IPNLDFFETYQADDFA YKAEGQGLAMEVDDLLFIQNYFKSIGCVPTETELKVLDTYWSDH 241

Query: 251 CRHTTFETELKNIDSASKFQKQLQATYDKYIAMRDELRSEKPQTLMMDMATIFGRYERA 310

CRHTTFETELKNIDSASKFQKQLQ TYDKYIAMRDELRSEKPQTLMMDMATIFGRYERA

Sbjct: 242 CRHTTFETELKNIDSASKFQKQLQTTYDKYIAMRDELRSEKPQTLMMDMATIFGRYERA 301

Query: 311 NGRLDDMEVSDEINACSVIEVDVGKVEPWLLMFKNETHNHPTEIEPFAGGAATCIGGAI 370

NGRLDDMEVSDEINACSVIEVDVGKVEPWLLMFKNETHNHPTEIEPFAGGAATCIGGAI

Sbjct: 302 NGRLDDMEVSDEINACSVIEVDVGKVEPWLLMFKNETHNHPTEIEPFAGGAATCIGGAI 361

45 Query: 371 RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY 430

RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY

Sbjct: 362 RDPLSGRSYVYQAMRISGAGDITTPIAETRAGKLPQQVISKTAAHGYSSYGNQIGLATTY 421

50 Query: 431 VREYFHPGFVAKRMELGAVVGAAPKENVVRKEPEAGDVVLLGGKTGRDGVGGATGSSKV 490

VREYFHPGFVAKRMELGAVVGAAPKENVVRKEPEAGDV+LLGGKTGRDGVGGATGSSKV

Sbjct: 422 VREYFHPGFVAKRMELGAVVGAAPKENVVRKEPEAGDVILLGGKTGRDGVGGATGSSKV 481

55 Query: 491 QTVESVETAGAEVQKGNAIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVVAIGELADGLE 550

QTVESVETAGAEVQKGNAIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVVAIGELADGLE

Sbjct: 482 QTVESVETAGAEVQKGNAIEERKIQRLFRDGNVTRLIKKSNDFGAGGVCVVAIGELADGLE 541

Query: 551 IDLDKVPLKYQGLNGTEIAISESESQERMSVVVGPSPDVDAFIAACNKENIDAVVVATVTEKP 610

IDLDKVPLKYQGLNGTEIAISESESQERMSVVV P+DVDAFIAACNKENIDAVVVATVTEKP

Sbjct: 542 IDLDKVPLKYQGLNGTEIAISESESQERMSVVVRPNVDADFIAACNKENIDAVVVATVTEKP 601

Query: 611 NLVMTWNGETIVDLERCFLDINGVRVVVDAKVVDKDLTVPEARTTSAETLEADMLKVLSD 670

NLVMTWNGE IVDLER FLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEAD LKVLSD

Sbjct: 602 NLVMTWNGEIIVDLERRFLDTNGVRVVVDAKVVDKDLTVPEARTTSAETLEADTLKVLSD 661

Query: 671 LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQYGVTITASVMAQGYN 730
 LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQ+GVTTTASVMAQGYN
 Sbjct: 662 LNHASQKGLQTIFDSSVGRSTVNHPIGGRYQITPTESSVQKLPVQHGVTITASVMAQGYN 721

5 Query: 731 PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQPVSALLG 790
 PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQPVSALLG
 Sbjct: 722 PYIAEWSPYHGAAYAVIEATARLVATGADWSRARFSYQEYFERMDKQAERFGQPVSALLG 781

10 Query: 791 STEAQIQFGLPSIGGKDMSMSGTFEEELTVPPTLVAFGTTADSRKVLSPEFKAAAGENIYYY 850
 STEAQIQ GLPSIGGKDMSMSGTFE+LTVPPTLVAFGTTADSRKVLSPEFKAAAGENIYYY
 Sbjct: 782 STEAQIQLGLPSIGGKDMSMSGTFEDELTVPPTLVAFGTTADSRKVLSPEFKAAAGENIYYY 841

15 Query: 851 PGQAISEDIDFDLIKANFSQFEAIQAOHKITAASAVKYGGVLESALMTFGNRIGASVEI 910
 PGQAISEDIDFDLIK NFSQFEAIQAOHKITAASA KYGGVLESALMTFGNRIGASVEI
 Sbjct: 842 PGQAISEDIDFDLIKDNFSQFEAIQAOHKITAASAALKYGGVLESALMTFGNRIGASVEI 901

20 Query: 911 AEELDSSLTAQLGGFVFTSVEEIADVVVKIGQTQADFTVTVNGNDLAGASLLSAFEKGKLEEV 970
 AEELDSSLTAQLGGFVFTS EEIAD VKIGQTQADFTVTVNGNDLAGASLL+AFEKGKLEEV
 Sbjct: 902 AEELDSSLTAQLGGFVFTSAEEIADAVKIGQTQADFTVTVNGNDLAGASLLAAFEKGKLEEV 961

25 Query: 971 YPTEFEQVDAIEEVPAVVDVVIKAKEIEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN 1030
 YPTEFEQ D +EVEPAVSD VIKAKE IEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN
 Sbjct: 962 YPTEFEQTDVLEEVPAVVDVTVIKAKETIEKPVVYIPVFPGTNSEYDSAKAFEQVGASVN 1021

30 Query: 1031 LVPFVTLNEAAIAESVDTMVANIKAIIFFAGGFSAADEPDGSAKFTVNILLNEKVRAA 1090
 LVPFVTLNE AIAESVDTMVANIKAIIFFAGGFSAADEPDGSAKFTVNILLNEKVRAA
 Sbjct: 1022 LVPFVTLNEVAIAESVDTMVANIKAIIFFAGGFSAADEPDGSAKFTVNILLNEKVRAA 1081

Query: 1091 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVERTRIA 1150
 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVERTRIA
 Sbjct: 1082 IDSFIEKGGLIIGICNGFQALVKSGLLPYGNFEEAGETSPTLFYNDANQHVAKMVERTRIA 1141

35 Query: 1151 NTNSPWLAGVEVGDIHVIPSHGEKGKFVVSASEFAELRDNGQIWSQYVDFDGQPSMDSKY 1210
 NTNSPWLAGVEVGDIH IPVSHGEKG VVSASEFAELRDNGQIWSQYVDFDGQPSMDSKY
 Sbjct: 1142 NTNSPWLAGVEVGDIHAIPVSHGEKGKLVVSASEFAELRDNGQIWSQYVDFDGQPSMDSKY 1201

Query: 1211 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIPGNQDKQLFESAVKYFTGK 1266
 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIPGNQDQ LF SAVKYFTGK
 Sbjct: 1202 NPNGSVNAIEGITSKNGQIIGKMGHSERWEDGLFQNIPGNQDKILFASAVKYFTGK 1257

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 305

A DNA sequence (GBSx0334) was identified in *S.agalactiae* <SEQ ID 983> which encodes the amino acid sequence <SEQ ID 984>. This protein is predicted to be phosphoribosylaminoimidazole-succinocarboxamide synthase (purC). Analysis of this protein sequence reveals the following:

Possible site: 41
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4783 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA03540 GB:L15190 SAICAR synthetase [Streptococcus pneumoniae]
 Identities = 183/231 (79%), Positives = 203/231 (87%)

60 Query: 1 MTNQLIYTGAKADITYSTKDENVIRTVYKDQATMLNGARKETIDGKGALNNQISSLIFEKL 60
 M+ QLIY+GAKADITY+T+DEN+I + YKDQAT NG +KE I GKG LNNQISS IFEKL
 Sbjct: 1 MSKQLIYSGKAKADITYTTEDENLIISTYKDQATAFNGVKKEQIAGKGVLNNQISSFIFEKL 60

-390-

Query: 61 NMAGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSF SKRFGVEEGH VLETPIVEFY 120
 N AGV TH++E++S EQLNKKV IIPLEVVLRN TAGSF SKRFGV+EG LETPIVEFY
 Sbjct: 61 NAAGVATHFVEKLS DTEQLNKKV KIIPLEVVLRN YTAGSF SKRFGV DEGIA LETPIVEFY 120

5

Query: 121 YKNDNLNDPFINDEHV KFLGIVNDEEIA YLKG ETRH INELLKD WFAQIGL NLID FKL EFG 180
 YKND+L+DPF INDEHV KFL I +D++IA YLK E R INELLK WFA+IGL LID FKL EFG
 Sbjct: 121 YKNDL DDPF INDEHV KFL QIADDQQI AYLKEEARR INELLK VWF A EIGL KL ID FKL EFG 180

10

Query: 181 FD KDGKII LADEFSPDN CRLWDADGNHMDK DVFR RD LGSL LTDV YQV VLEKL 231
 FD KDGKII LADEFSPDN CRLWDADGNHMDK DVFR LG LTDV Y++V EKL
 Sbjct: 181 FD KDGKII LADEFSPDN CRLWDADGNHMDK DVFR RLGE LT D VYEIVWEKL 231

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 985> which encodes the amino acid sequence <SEQ ID 986>. Analysis of this protein sequence reveals the following:

Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 20 bacterial cytoplasm --- Certainty=0.3935 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

25 Identities = 221/234 (94%), Positives = 228/234 (96%)

Query: 1 MTNQLI YTGKAKDIYSTKDEN VIRT VYKDQATMLNGARKETIDGKG ALNNQISSLIFEKL 60
 +TNQLI Y GKAKDIYSTKDEN VIRT VYKDQATMLNGARKETIDGKG ALNNQISSLIFEKL
 Sbjct: 11 VTNQLI YKGKAKDIYSTKDEN VIRT VYKDQATMLNGARKETIDGKG ALNNQISSLIFEKL 70

30 Query: 61 NMAGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSF SKRFGVEEGH VLETPIVEFY 120
 N AGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSF SKRFGVEEGH VLETPIVEFY
 Sbjct: 71 NKAGVVTHYIEQISKNEQLNKKVDIIPLEVVLRNVTAGSF SKRFGVEEGH VLETPIVEFY 130

35 Query: 121 YKNDNLNDPFINDEHV KFLGIVNDEEIA YLKG ETRH INELLKD WFAQIGL NLID FKL EFG 180
 YKND+L+DPF INDEHV KFL I +D++IA YLK E R INELLK WFA+IGL LID FKL EFG
 Sbjct: 131 YKNDL DDPF INDEHV KFLGIVNDEEIA YLKG ETRR INELLKG WFAQIGL NLID FKL EFG 190

40 Query: 181 FD KDGKII LADEFSPDN CRLWDADGNHMDK DVFR RD LGSL LTDV YQV VLEKL JAL 234
 FD++G IIIA DEFSPDN CRLWD +GNHMDK DVFR RD LG+LTDV YQV VLEKL JAL
 Sbjct: 191 FD QEGT IIIA DEFSPDN CRLWDKG NHMDK DVFR RD LGNL LTDV YQV VLEKL JAL 244

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 306

A DNA sequence (GBSx0335) was identified in *S.agalactiae* <SEQ ID 987> which encodes the amino acid sequence <SEQ ID 988>. Analysis of this protein sequence reveals the following:

Possible site: 13
 >>> Seems to have no N-terminal signal sequence

50 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2779 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

55 A related GBS nucleic acid sequence <SEQ ID 9457> which encodes amino acid sequence <SEQ ID 9458> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

-391-

>GP:AAC35700 GB:AF041468 acyl carrier protein [Guillardia theta]
Identities = 27/75 (36%), Positives = 52/75 (69%)

5 Query: 12 MSRDEVFEKMLELLRQQQLGDPQLDITPESSLHDDLAIDSIALTEFIINLEDVFHLEIPDE 71
M+ E+FEK+ ++ +QLG + +T +++ +DL DS+ E ++ +E+ F++EIPD+
Sbjct: 1 MNEQEIEFEKVQTIISEQLGVDKSQVTKDANFANDLGADSLDTVELVMAIEAFNIEIPDD 60

Query: 72 AVEHMSSVQQQLLDYI 86
A E +S++QQ +D+I
10 Sbjct: 61 AAEQISNLQQAVDFI 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 989> which encodes the amino acid sequence <SEQ ID 990>. Analysis of this protein sequence reveals the following:

Possible site: 24
15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
bacterial cytoplasm --- Certainty=0.1917(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
20 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 36/77 (46%), Positives = 57/77 (73%)

25 Query: 12 MSRDEVFEKMLELLRQQQLGDPQLDITPESSLHDDLAIDSIALTEFIINLEDVFHLEIPDE 71
M+R E+FE+++ L+++Q + IT ++ L +DLA+DSI L EFIIN+ED FH+ IPDE
Sbjct: 1 MTRQEIFERLINLIQKQRSYLSVAITEQTHLKNDLAVDSIELVEFIINVEDFHIAIPDE 60

Query: 72 AVEHMSSVQQQLLDYIIE 88
VE M ++ +LDY+++
30 Sbjct: 61 DVEDMVFMRDILDYLVQ 77

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

35 Example 307

A DNA sequence (GBSx0336) was identified in *S.agalactiae* <SEQ ID 991> which encodes the amino acid sequence <SEQ ID 992>. This protein is predicted to be fatty acid/phospholipid synthesis protein (plsX). Analysis of this protein sequence reveals the following:

Possible site: 21
40 >>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -0.64 Transmembrane 101 - 117 (101 - 117)

----- Final Results -----
bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
45 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9455> which encodes amino acid sequence <SEQ ID 9456> was also identified.

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13462 GB:Z99112 alternate gene name: ylpD [Bacillus subtilis]
Identities = 174/329 (52%), Positives = 238/329 (71%), Gaps = 2/329 (0%)

55 Query: 8 KIAIDAMGGDYAPKAIVEGVNQAIISDFSDIEVQLYGDQKKIEKYLTVT-ERVSIIHTEEK 66
+IA+DAMGGD+APKA+++GV + I F D+ + L GD+ IE +LT T +R++++H +E
Sbjct: 2 RIAVDAMGGDHAPKAVIDGVIKGIEAFDDLHITLVDKTTIESHLTTSDRITVLHADEV 61

-392-

Query: 67 INSDDEPAKAVRRKKQSSMVLGAKAVKDGAQAFISAGNTGALLAAGLFVVGRIKGVD 126
 I DEP +AVRRKK SSMVL A+ V + A A ISAGNTGAL+ AGLF+VGRIKG+DRP
 Sbjct: 62 IEPTDEPVRAVRRKKNSSMVLMAQEVAENRADACISAGNTGALMTAGLFIVGRIKGIDRP 121

5 Query: 127 GLMSTMPTLDGVGFDMLDLGANAENTASHLHQYAILGSFYAKNVRGIEPVGLLNNGT 186
 L T+PT+ G GF +LD+GAN + HL QYAI+GS Y++ VRG+ PRVGLLN GTE
 Sbjct: 122 ALAPTLPTVSGDGFLLLDVGANVDAKPEHLVQYAIMGSVVSQQRGVTSPrVGLLNVGT 181

10 Query: 187 ETKGDSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVTDGFTGNAVLKTMEGTAMS 246
 + KG+ L K+ ++L +INFIGN+EARDL+ VADVVTDGFTGN LKT+EG+A+S
 Sbjct: 182 DKKGNELTKQTFOILKETANINFIGNVEARDLLDDVADVVTDGFTGNVTLKTLEGSALS 241

15 Query: 247 IMGSLKSSIKGGSVKAALKGALLKDSLYQLKDSMDYSSAGGAVLFGLKAPIVKCHGSSDS 306
 I ++ + + + +KL A +LK L ++K M+YS+ GGA LFGLKAP++K HGSSDS
 Sbjct: 242 IFKMMR-DVMTSTLTSKLAALVLPKPLKEMKMKMEYSNYGGASLFGLKAPVIKAHGSSDS 300

20 Query: 307 KAVYSTLKQVRTMLETQVVDQLVDAFTDE 335
 AV+ ++Q R M+ V + + +E
 Sbjct: 301 NAVFHAIRQAREMVSQNVAALIQQEVKEE 329

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 993> which encodes the amino acid sequence <SEQ ID 994>. Analysis of this protein sequence reveals the following:

Possible site: 36
 25 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.07 Transmembrane 121 - 137 (120 - 138)

----- Final Results -----
 bacterial membrane --- Certainty=0.1829(Affirmative) < succ>
 30 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9127> which encodes the amino acid sequence <SEQ ID 9128>. Analysis of this protein sequence reveals the following:

35 Possible cleavage site: 16
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -2.07 Transmembrane 95 - 111 (94 - 112)

40 ----- Final Results -----
 bacterial membrane --- Certainty= 0.183(Affirmative) < succ>
 bacterial outside --- Certainty= 0.000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

45 Identities = 254/330 (76%), Positives = 290/330 (86%)

Query: 6 MKKIAIDAMGGDYAPKAIVEGVNQAISDFSDIEVQLYGDQKKIEKYLTVERSVIIHTEE 65
 MK+IAIDAMGGD APKAIVEGVNQAI FSDIE+QLYGDQ KI YL ++RV+IIHT+E
 Sbjct: 27 MKRIAIDAMGGDNAPKAIVEGVNQAIASFSDIEIQLYGDQTKINSYLIQSDRVAAIHTE 86

50 Query: 66 KINSDEPAKAVRRKKQSSMVLGAKAVKDGAQAFISAGNTGALLAAGLFVVGRIKGVD 125
 KI SDDEPAKAVRKK++SMVL AKAVK+G A A ISAGNTGALLA GLFVVGRIKGVD
 Sbjct: 87 KIMSDDEPAKAVRKKKASMVLAAKAVKEGKADAIISAGNTGALLAVGLFVVGRIKGVD 146

55 Query: 126 PGLMSTMPTLDGVGFDMLDLGANAENTASHLHQYAILGSFYAKNVRGIEPVGLLNNGT 185
 PGL+ST+PT+ G+GFDMLDLGANAENTA HLHQYAILGSFYAKNVRGIEPVGLLNNGT
 Sbjct: 147 PGLLSTIPTVTGLGFDMLDLGANAENTAKHLHQYAILGSFYAKNVRGIANPRVGLLNNGT 206

60 Query: 186 EETKGDSLHKEAYELLAAEPSINFIGNIEARDLMSSVADVVTDGFTGNAVLKTMEGTAM 245
 EETKGD L K YELL A+ +I+F+GN+EAR+LMS VADV+V+DGFTGNAVLK++EGTA+
 Sbjct: 207 EETKGDPLRKATYELLTADNTISFVGNVEARELMSGVADVIVSDGFTGNAVLKSIEGTAI 266

Query: 246 SIM GSLKSSIKGGSVKAALKGALLKDSLYQLKDSMDYSSAGGAVLFGLKAPIVKCHGSSD 305

-393-

SIMG LK I SGG+K K+GA LLK SLY++K ++DYSSAGGAVLFGLKAP+VK HGSSD
 Sbjct: 267 SIMGQLKQIINSGGIKTKIGASLLKSSLYEMKKTLDYSSAGGAVLFGLKAPVVKSHGSSD 326

Query: 306 SKAVYSTLKVQVRTMLETQVVDQLVDAFTDE 335
 KA++ST+KQVRTML+T VV QLV+ F E
 Sbjct: 327 VKAIFSTIKQVVRTMLDTNVVGQLVEFAKE 356

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 308

A DNA sequence (GBSx0337) was identified in *S.agalactiae* <SEQ ID 995> which encodes the amino acid sequence <SEQ ID 996>. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4668 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 309

A DNA sequence (GBSx0338) was identified in *S.agalactiae* <SEQ ID 997> which encodes the amino acid sequence <SEQ ID 998>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

30 INTEGRAL Likelihood = -12.84 Transmembrane 61 - 77 (55 - 82)
 INTEGRAL Likelihood = -10.14 Transmembrane 26 - 42 (19 - 51)
 INTEGRAL Likelihood = -9.77 Transmembrane 192 - 208 (186 - 211)
 INTEGRAL Likelihood = -5.79 Transmembrane 267 - 283 (262 - 286)
 INTEGRAL Likelihood = -3.77 Transmembrane 100 - 116 (99 - 116)

35 ----- Final Results -----
 bacterial membrane --- Certainty=0.6137 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

40

A related GBS nucleic acid sequence <SEQ ID 9453> which encodes amino acid sequence <SEQ ID 9454> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA22372 GB:AL034446 putative transmembrane protein
 [Streptomyces coelicolor A3(2)]
 Identities = 47/154 (30%), Positives = 69/154 (44%), Gaps = 12/154 (7%)

45

Query: 120 SGFVEISSLNSFSFGPFFFFLFLAYFIQLSLTEELFRGYVMTTFTKFKGSFAGVLCNSMLF 179
 SG+ E+ S F+A + TEE++FRG + + + G++ + ++F
 Sbjct: 118 SGYYEVGLGSVQGAIGLVGFMA---AAAATEEVVFRGVLFRIIEEHIGTYLALGLTGLVF 175

Query: 180 SFIHFRN-----YGITAIALFNLFLGIIFSILFNMTKNILFVTGVHTTWNTMGCVLGN 234
 +H N +G AIA+ F+L ++ T+N+ GVH WNF G V

-394-

Sbjct: 176 GLMHLLNEDATLWGALAIATEAGFMLAAAYAA---TRNLWLTIJVHFGWNFAAGGVFST 231

Query: 235 KVSGGDSPVSLFRITENSSFALWNGGDFGFEGGV 268

VSG L T S L GGDFG EG V

5 Sbjct: 232 VVSGNGDSEGLLDAT-MSGPKLLTGGDFGPEGSV 264

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 310

A DNA sequence (GBSx0339) was identified in *S.agalactiae* <SEQ ID 999> which encodes the amino acid sequence <SEQ ID 1000>. Analysis of this protein sequence reveals the following:

Possible site: 55

>>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2665 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20 A related GBS nucleic acid sequence <SEQ ID 9451> which encodes amino acid sequence <SEQ ID 9452>

was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05088 GB:AP001511 unknown conserved protein [Bacillus halodurans]
25 Identities = 81/242 (33%), Positives = 124/242 (50%), Gaps = 3/242 (1%)

Query: 8 GLVLYNRNYREDDKLVKIIFTETEGKRMFFVKHAS--KSKFNAVLQPLTIAHFILKINDNG 65
G+V+ +Y E +K+V +FT GK + A KS+ AV Q T + + N G

Sbjct: 7 GIVIRTVVDYGESNKIVTVFTREYGKIALMARGAKRPKSRLTAVTQLFTYGMFFFQKNA-G 65

30 Query: 66 LSYIDDYKEVLAFQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTLELIEDG 125
L + + + +F+E +DLF+ SY SY+T L + D + LF L +T+ + +G
Sbjct: 66 LGTLTQGEIIQSFRREVNRDLFRASYVSYVTDLTNKLTEDEKRNPYLFELLYQTIHYMNEG 125

35 Query: 126 LDYEILTNIFEVQLLERFGVALNFHDCCVFCHRVGLPFDFSHKYSGLLCPNHYYKDERRNH 185
+D ++LT IFEV++ G+ CV C +P FS K +G LC KD
Sbjct: 126 MDPDVLTTRIFEVKMFTVAGIKPELDQCVSCRSTDVPVGFSIKEAGFLCKRCIEKDPMAYK 185

40 Query: 186 LDPNMLYLINRFQSIQFDDLQITISVKPEMKLKIROFLDMIYDEYVGIGHLKSKKFIDDLSSWG 247
+ + L+ F L TIS+KPE K ++ + YDEY G+HLKS++F+D L S G
Sbjct: 186 ITAQVAKLLRLFYHFDLQRLGTISLKPETKATLKTIIHQYYDEYSGLHLKSRRFLDQLESMG 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1001> which encodes the amino acid sequence <SEQ ID 1002>. Analysis of this protein sequence reveals the following:

45 Possible site: 46

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1566 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 159/251 (63%), Positives = 210/251 (83%)

55 Query: 1 MRVSQTYGLVLYNRYREDDKLVKIIFTETEGKRMFFVKHASKSKFNAVLQPLTIAHFILK 60

-395-

```

M+++++ G+VL+NRNYREDDKLVKIFTE GK+MFFVKH S+SK +++QPLTIA FI K
Sbjct: 1 MQLTESLGIVLFNRNYREDDKLVKIFTEVAGKQMFFVKHISRSKMSSIIQPLTIADFIK 60

5 Query: 61 INDNGLSYIDDYKEVLAQETNSDLFKLSYASYITSLADVAISDNVADAQLFIFLKKTLE 120
      +ND GLSY+ DY V ++ N+D+F+L+YASY+ +LAD AI+DN +D+ LF FLKKTLE+
Sbjct: 61 LNDTGLSYVVVDYSNVNTYRYINNDIFRLAYASYVLALADAIAADNESDSHLFTFLKKTLD 120

Query: 121 LIEDGLDYEILTNIIFEVQLLERFGVALNFHDGVFCRVGLPFDHSKYSGLLCPNHYKD 180
      L+E+GLDYEILTNIIFE+Q+L+RFG++LNFH+C CHR LP DFSH++S +LC HYKD
10 Sbjct: 121 LMEEGLDYEILTNIIFEIQILDRCFGISLNHFECAIChRTDLPLDFSHRFSAVLCSEHYKD 180

Query: 181 ERRNHLDPNMLYLINRFQSIQFDDLQTISVKPEMKLKIRQFLDMIYDEYVGHLKSKKFI 240
      RRRNHLDPN++YL++RFQ I FDDL+TIS+ ++K K+RQF+D +Y +YVG I LSKK FI
Sbjct: 181 NRRRNHLDPNVIYLLSRFQKITFDDLRTISLNKDIKKLRQFIDELYHDYVGIKLKSCKFI 240

15 Query: 241 DDLSSWGSIMK 251
      D+L WG IMK
Sbjct: 241 DNLVKWDIMK 251

```

20 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 311

25 A DNA sequence (GBSx0340) was identified in *S.agalactiae* <SEQ ID 1003> which encodes the amino acid sequence <SEQ ID 1004>. This protein is predicted to be aromatic amino acid aminotransferase (patA). Analysis of this protein sequence reveals the following:

```

Possible site: 14
>>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -3.13 Transmembrane 141 - 157 ( 140 - 159)

30 ----- Final Results -----
      bacterial membrane --- Certainty=0.2253 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

35 A related GBS nucleic acid sequence <SEQ ID 9449> which encodes amino acid sequence <SEQ ID 9450> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAF06954 GB:AF146529 aromatic amino acid aminotransferase
      [Lactococcus lactis subsp. cremoris]
40 Identities = 261/391 (66%), Positives = 323/391 (81%)

      Query: 38 MTLEKRFNKYLDRIEVSLIRQFDQSISDIPGMVKLTGEPDFTPDHVKEAKSAIDANQ 97
              M L K+FN LD+IE+SLIRQFDQ +S IP ++KLTLGEPDF TP+HVK+A +AI+ NQ
      Sbjct: 1 MDLLKKFNPNLKIEISLIRQFDQQVSSIPDIILTLGEPDFYTPHEHVQAGIAAIENNQ 60

45 Query: 98 SYYTGMMSGLLALRQAAADFAKDKYNLTYNPDCEILVTIGATEALASASLIAILEAGDVILL 157
      S+YTGM+GLL LRQAA++F KY L+Y + EILVT+G TEA+S+ L++IL AGD VL+
      Sbjct: 61 SHYTGMAGLLELRQAASEFLKKYGLSYAAEDEILTVGVTEAISSVLLSILVAGDEVLI 120

50 Query: 158 PAPAYPGYEPIVNLVGADIVEIITRENDFRITPEMLETAIQQGEKLKAVILLNYPTNPTG 217
      PAPAYPGYEP++ L G +VEIDTR NDF LTPEML+ AII++ K+KAV+LNYP NPTG
      Sbjct: 121 PAPAYPGYEPITLAGGSIVEIDTRANDFVLTPEMLDQAIIEREGKVKAVILNYPANPTG 180

55 Query: 218 ITYSRQEJAALAEVLKKYDIFVISDEVYSELTYTGQQHVSIAEYLPNQQTILINGLSKSHA 277
      +TY+R++I LAEVLKK+++FVI+DEVYSEL YT Q HVSIAEY P QTI++NGLSKSHA
      Sbjct: 181 VTYNREQIKDLAEVLKKHEVFVIADEVYSELNYTDQPHVSTAEYAPEQTIVLNGLSKSHA 240

Query: 278 MTGWRVGLVYAPEAFIAQIIKSHQYMVTAASTISQFAGVEALSVGKNDTLPMRQGYIKRR 337
      MTGWR+GL++A +AQIIK+HQY+VT+AST SQFA +EAL G +D LPM++ Y+KRR

```

-396-

Sbjct: 241 MTGWRIGLIFAARELVAQI IKTHQYLVTSASTQSQFAAIEALKNGADDALPMKKEYLKRR 300

Query: 338 DYIIDKMSKLGFKIIKPSGAFYIFAKIPDSYPQDSFKFCQDFAYQQAVAIIPGVAFGKYG 397
DYI+KMS LGFKII+P GAFYIFAKIP QDSFKF DFA + AVAIIPG+AFG+YG

5 Sbjct: 301 DYIIEKMSALGFKIIIEPDGAFYIFAKIPADLEQDSFKFAVDFAKENAVAIIPGIAFGQQG 360

Query: 398 EGYIRLSYAASMEVIETAMARLKVFMESYEG 428
EG++RLSYAASM+VIE AMARL ++ G

Sbjct: 361 EGFVRLSYAASMDVIEQAMARLTDYVTKKRG 391

10

There is also homology to SEQ ID 1006.

SEQ ID 1004 (GBS332) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 60 (lane 3; MW 50.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 4; MW 76kDa).

15

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 312

A DNA sequence (GBSx0341) was identified in *S.agalactiae* <SEQ ID 1007> which encodes the amino acid sequence <SEQ ID 1008>. This protein is predicted to be ribose-phosphate pyrophosphokinase (prsA).

20

Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence
```

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.3118(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

25

A related GBS nucleic acid sequence <SEQ ID 9447> which encodes amino acid sequence <SEQ ID 9448> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA62181 GB:M92842 prs [Listeria monocytogenes]
  Identities = 209/312 (66%), Positives = 266/312 (84%), Gaps = 3/312 (0%)
```

35

```
Query: 10 LKLFALSSNKELAKKVSQLTIGIPLGQSTVRQFSDGEIQVNIEESIRGHHVFIQLQSTSPPV 69
      LK+F+L+SN+ELA+++++ +GI LG+S+V FSDGEIQ+NIEESIRG HV+++QSTS+PV
Sbjct: 10 LKIFSLSNSNRELAAEIAKEVGVIELGKSSVTHFSDGEIQINIEESIRGCHVYVIQSTSNPV 69
```

40

```
Query: 70 NDNLMEILIMVDALKRASAESVSVMPYYGYARQDRKARSREPITSKLVANMLEVAGVDR 129
      N NLME+LIM+DALKRASA +++, VMPYYGYARQDRKARSREPIT+KLVAN+E AG R
Sbjct: 70 NQNLMELLIMIDALKRASAATINTIVMPYYGYARQDRKARSREPITAKLVANLIETAGATR 129
```

45

```
Query: 130 LLTVIDLHAAQIQQFFDIPVVDHLMGAPLIADYFDRQGLVGGDVVVSPDHGGVTRARKLAQ 189
      ++T+D+HA QIQGFFDIP+DHL L++DYF + L GDD+VVVSPDHGGVTRARK+A
Sbjct: 130 MITLDMHAPQIQQFFDIPIDHNAVRLLSDYFSERHL-GDDLVVVSPDHGGVTRARKMAD 188
```

50

```
Query: 190 CLKTPIAIIDKRRSVTKMNTSEVMNIIGNIKGKKCILIDDMIDTAGTICHAADALAEAGA 249
      LK PIAIIDKRR + N +EVMNI+GN++GK CI+IDD+IDTAGTI AA AL EAGA
Sbjct: 189 RLKAPIAIIDKRR--PRPNVAEVNMIVGNVEGKVCIIIDDIIDTAGTITLAAKALREAGA 246
```

55

```
Query: 250 TAVYASCTHPVLSGPALDNIONSAIEKLVLDTIYLPEERLIDKIEQISIAELIGEAIIR 309
      T VYA C+HPVLSGP+ I+ S IEKL+V ++I LPEE+ IDK+EQ+S+A L+GEAI+R
Sbjct: 247 TKVYACCSHPVLSGPAMKRIEESPIEKLVVTNSIALPEEKWIDKMEQLSVAALLGEAIVR 306
```

55

```
Query: 310 IHEKRPLSPLFE 321
      +HE +S LFE
```

-397-

Sbjct: 307 VHENASVSSLFE 318

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1009> which encodes the amino acid sequence <SEQ ID 1010>. Analysis of this protein sequence reveals the following:

5 Possible site: 22
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.2685 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

15 Identities = 298/322 (92%), Positives = 311/322 (96%)
 Query: 1 MEEIMSYSNLKLFALSSNKELAKKVQSTIGIPLGQSTVRQFSDGEIQVNIEESIRGHVF 60
 +EE MSYS+LKLFA LSSNKELA+KV+ +GI LG+STVRQFSDGEIQVNIEESIRGHVF
 Sbjct: 1 LEEKMSYSDLKLFALSSNKELAEKVASAMGIQLGKSTVRQFSDGEIQVNIEESIRGHVF 60
 20 Query: 61 ILQSTSSPVNDNLMEILIMVDALKRASAESVVMPYYGYARQDRKARSREPITSKLVAN 120
 ILQSTSSPVNDNLMEILIMVDALKRASA +SVVMPYYGYARQDRKARSREPITSKLVAN
 Sbjct: 61 ILQSTSSPVNDNLMEILIMVDALKRASAEKISVVMPYYGYARQDRKARSREPITSKLVAN 120
 25 Query: 121 MLEVAGVDRLLTVDLHAAQIQQFFDIPVDHLMGAPIIADYFDRQGLVGDDVVVVSPDHGG 180
 MLEVAGVDRLLTVDLHAAQIQQFFDIPVDHLMGAPIIADYFDR GLVG+DVVVVSPDHGG
 Sbjct: 121 MLEVAGVDRLLTVDLHAAQIQQFFDIPVDHLMGAPIIADYFDRHGLVGEDVVVVSPDHGG 180
 30 Query: 181 VTRARKLAQCLKTPIAIIDKRRSVTKMNTSEVMNIIGNIKGKKCILIDDMIDTAGTICHA 240
 VTRARKLAQ L+TPIAIIDKRRSV KMNTSEVMNIIGN+ GKKCILIDDMIDTAGTICHA
 Sbjct: 181 VTRARKLAQFLQTPIAIIDKRRSVDKMNTSEVMNIIGNVSGKKCILIDDMIDTAGTICHA 240
 35 Query: 241 ADALAEAGATAVYASCTHPVLSGPALDNIQNSAIEKLIVLDIYLPEERLIDKIEQISIA 300
 ADALAEAGATAVYASCTHPVLSGPALDNIQ SAIEKLIVLDIYLPEERLIDKIEQISIA
 Sbjct: 241 ADALAEAGATAVYASCTHPVLSGPALDNIQRSAIEKLIVLDIYLPEERLIDKIEQISIA 300
 Query: 301 ELIGEAIIRIHEKRPLSPLFEM 322
 +L+ EAIIRIHEKRPLSPLFEM
 Sbjct: 301 DLVAEAIIRIHEKRPLSPLFEM 322

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 313

A DNA sequence (GBSx0342) was identified in *S.agalactiae* <SEQ ID 1011> which encodes the amino acid sequence <SEQ ID 1012>. This protein is predicted to be a secreted protein. Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.3751 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9277> which encodes amino acid sequence <SEQ ID 9278> 55 was also identified.

The protein has homology with the following sequences in the GENPEPT database:

-398-

>GP:AAD00288 GB:U78607 putative secreted protein [Streptococcus mutans]
 Identities = 111/157 (70%), Positives = 130/157 (82%), Gaps = 1/157 (0%)

5 Query: 1 MTAIKGVGALESQQSELEAQNAQLEAVSQQLGQEIQTLSNKIVARNESLKKQVRSAQKG 60
 + I+GQV AL++QQ+EL+A+N +LEA S LGQ+IQTLS+KIVARNESLK+Q RSAQK
 Sbjct: 55 LITIQGVSAQIQTQQAELQQAENQRLEAQSATLGQQIQTLSKIVARNESLKQQQARSAQKS 114

10 Query: 61 NL-TNYINTILNSKSVSDAVNRVVAIREVVVSANEKMLAQOEADKAALAKQIENQNAINT 119
 N T+YIN I+NSKSVSDA+NRV AIREVVVSANEKML QOE DKAA+E KQ ENQ AINT
 Sbjct: 115 NAATSYINAIINSKSVSDAINRVSIAIREVVVSANEKMLQQQEVDKAAVEQKQQENQAAINT 174

15 Query: 120 VAANKQAIENNKAALATQRAQLEAAQLELSAQLTTVQ 156
 VAAN++ I N AL TQ+AQLEAAQL L A+LTQ Q
 Sbjct: 175 VAANQETIAQNTNALNTQQAQLEAAQLNQAELETTAQ 211

There is also homology to SEQ ID 1014.

A related GBS gene <SEQ ID 8543> and protein <SEQ ID 8544> were also identified. Analysis of this protein sequence reveals the following:

20 Lipop: Possible site: -1 Crend: 3
 McG: Discrim Score: 8.29
 GvH: Signal Score (-7.5): 0.8
 Possible site: 49
 >>> Seems to have a cleavable N-term signal seq.
 ALOM program count: 0 value: 6.74 threshold: 0.0
 25 PERIPHERAL Likelihood = 6.74 400
 modified ALOM score: -1.85

*** Reasoning Step: 3

30 ----- Final Results -----
 bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear)

35 The protein has homology with the following sequences in the databases:

32.8/56.3% over 439aa

Lactococcus lactis

40 GP|512521| usp 45 Insert characterized
 PIR|JN0097|JN0097 secreted 45K protein precursor - Insert characterized

ORF00094 (301 - 1563 of 1941)
 45 GP|512521|emb|CAA01320.1||A17083(1 - 440 of 461) usp 45 {Lactococcus
 lactis}PIR|JN0097|JN0097 secrete
 d 45K protein precursor - Lactococcus lactis
 %Match = 16.5
 %Identity = 32.8 %Similarity = 56.3
 Matches = 141 Mismatches = 178 Conservative Sub.s = 101

50 93 123 153 183 213 243 273 303
 RKYYNFKNSTLFLFLF*FHYGVIIIE*IEEGYRFLDLIMVHLETVDKYKCNDVI*FREFFGKIFNVLS*RSSLIKM

|
M

55 333 387 417 447 477 507 537
 KKRILSAVLVSGVTLGTAA--VTVNADDFDKIAATDSVINLTSQQAAAQNQVTAIKGVGALESQQSELEAQNAQLEA
 ||:||:||:|| | | || :| || | :|::| || || ::| :| :| :| :| :| :| :| :| :| :| :| :| :|
 KKKIISAILMSTVILSAAAPLSGVYAD-TNSDIAKQDATISSAQSAKAQAOQVDSLQSKVDSLQQKQTSTKAQIAKIES

20 30 40 50 60 70 80

60 567 597 627 654 684 714 744 774
 VSQQQLGQEIQTLSNKIVARNESLKKQVRSAQ-KGNLTNYINTILNSKSVSDAVNRVVAIREVVVSANEKMLAQOEADKAAL
 ::| :| ||:| | | ::|:| || || :| ||::| ||::| :| :| | |||:| || || ::| |
 EAKALNAQIATLNESIKERTKTLEAQARSAQVNSSATNYMDAVVNSKSLTDVIQKVTIAITVSSANKQMLEQQEKEQKEL

90 100 110 120 130 140 150 160

SEQ ID 8544 (GBS65) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 5 (lane 6; MW 47.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 3; MW 72kDa) and in Figure 175 (lane 2 & 3; MW 72kDa).

30 The GBS65-GST fusion product was purified (Figure 102A; see also Figure 191, lane 4) and used to immunise mice (lane 1 product; 20 μ g/mouse). The resulting antiserum was used for Western blot (Figure 102B), FACS, and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 314

A DNA sequence (GBSx0343) was identified in *S.agalactiae* <SEQ ID 1015> which encodes the amino acid sequence <SEQ ID 1016>. Analysis of this protein sequence reveals the following:

Possible site: 18
40 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1184(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

45 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in

No corresponding DNA sequence was identified in *S.pyogenes*.
Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
50 vaccines or diagnostics.

-400-

Example 315

A DNA sequence (GBSx0344) was identified in *S.agalactiae* <SEQ ID 1017> which encodes the amino acid sequence <SEQ ID 1018>. Analysis of this protein sequence reveals the following:

5 Possible site: 23
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4736 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 10 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for
 15 vaccines or diagnostics.

Example 316

A DNA sequence (GBSx0345) was identified in *S.agalactiae* <SEQ ID 1019> which encodes the amino acid sequence <SEQ ID 1020>. This protein is predicted to be elongation factor Tu (tufA). Analysis of this protein sequence reveals the following:

20 Possible site: 43
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3012 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9737> which encodes amino acid sequence <SEQ ID 9738> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB03851 GB:AP001507 translation elongation factor Tu (EF-Tu)
 [Bacillus halodurans]
 Identities = 302/397 (76%), Positives = 350/397 (88%), Gaps = 2/397 (0%)

35 Query: 7 MAKEKYDRSKPHVNIGTIGHVDHGKTTLTAITTTLARRLPTSVNQPKDYASIDAPEER 66
 MAKEK+DRSK H NIGTIGHVDHGKTTLTAITTTLA+R V Y +ID APEER
 Sbjct: 1 MAKEKFDRSKTHANIGTIGHVDHGKTTLTAITTVLAKRSGKGVAMA--YDAIDGAPEER 58

40 Query: 67 ERGITINTAHVEYETEKRHAYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDGMPQTR 126
 ERGIDI+TAHVEYET+ RHYAH+D PGHADYVKNMITGAAQMDG ILVV++ DGMPQTR
 Sbjct: 59 ERGITISTAHVEYETDNRHYAHVDCPGHADYVKNMITGAAQMDGGILVVSAADGMPQTR 118

Query: 127 EHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEIMEIRDLLSEYDFPGDDLPVIQGSALK 186
 EHILLSRQVGV +L+VF+NK D+VDDEELLELVEIME+RDLLSEYDFPGDD+PVI+GSALK
 45 Sbjct: 119 EHILLSRQVGVPYLVVFLNKCDMVDDDEELLELVEIMEVRDLLSEYDFPGDDVPVIRGSALK 178

Query: 187 ALEGDEKYEDIIMELMSTVDEYIPEPERDTDKPLLLPVEDVFSITGRGTVASGRIDRGTV 246
 ALEGD ++E+ I+ELM+ VD+YIP PERDT+KP ++PVEDVFSITGRGTVA+GR++RG +
 Sbjct: 179 ALEGDAEWEKKIELMAAVDDYIPTPERDTKEPKFMMMPVEDVFSITGRGTVATGRVERGQL 238

50 Query: 247 RVNDEVEIIVGIKEDIQKAVVTGVEMFRKQLDEGLAGDNVGVLRLCVQRDEIERGQVLAKP 306
 V DEVEI+G++E+ +K VTGVEMFRK LD AGDN+G LLRGV R+E++RGQVLAKP
 Sbjct: 239 NVGDEVEIIGLEEAKTTVTGVEMFRKLLDYAEAGDNIGALLRGVSREEVQRGQVLAKP 298

-401-

Query: 307 GSINPHTRFKGEVYIILSKEEGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEMVMPGDN 366
 G+I PHT FK EVY+LSKEEGGRHTPFF+NYRPQFYFRTTDVTG I+LP G EMVMPGDN
 Sbjct: 299 GTITPHTNFKAEVYVLSKEEGGRHTPFFSNYRPQFYFRTTDVTGIIQLPDGVEMVMPGDN 358

5 Query: 367 VTIEVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIE 403
 V + VELI PIA+E+GT FSIREGGRTVG+G+V+ I+
 Sbjct: 359 VEMTVELIAPIAEETKFSIREGGRTVGAGVVASIQ 395

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1021> which encodes the amino acid
 10 sequence <SEQ ID 1022>. Analysis of this protein sequence reveals the following:

Possible site: 19
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.1367 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

20 Identities = 386/404 (95%), Positives = 396/404 (97%)

Query: 1 MEAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTLAAITTVLARRLPTSVNQPKDYASID 60

+EAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTLAAITTVLARRL+SVNQPKDYASID

25 Sbjct: 12 LEAFPKMAKEKYDRSKPHVNIGTIGHVDHGKTTLAAITTVLARRLPSSVNQPKDYASID 71

Query: 61 AAPEERERGITINTAHVEYETEKRHAYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 120

AAPEERERGITINTAHVEYET RHAYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG

Sbjct: 72 AAPEERERGITINTAHVEYETATRHAYAHIDAPGHADYVKNMITGAAQMDGAILVVASTDG 131

30 Query: 121 PMPQTREHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPVI 180
 PMPQTREHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPVI

Sbjct: 132 PMPQTREHILLSRQVGVKHLIVFMNKVDLVDDEELLELVEMEIRDLLSEYDFPGDDLPVI 191

35 Query: 181 QGSALKALEGDEKYEDIIMELMSTVDEYIYPEPERDTDKPLLLPVEDVFSITGRGTVASGR 240
 QGSALKALEGD K+EDIIMELM TVD YIYPEPERDTDKPLLLPVEDVFSITGRGTVASGR
 Sbjct: 192 QGSALKALEGDTKFEDIIMELMDTVDSYIYPEPERDTDKPLLLPVEDVFSITGRGTVASGR 251

40 Query: 241 IDRGTVRVNDEVEIVGIKEKEDIQAKAVTGEMFRKQLDEGLAGDNVGVLRLRGVQRDEIERG 300
 IDRGTVRVNDE+EIVGIKE+ +KAVTGEMFRKQLDEGLAGDNVG+LLRGVQRDEIERG
 Sbjct: 252 IDRGTVRVNDEIEIVGIKEETKKAVTGEMFRKQLDEGLAGDNVGILLRGVQRDEIERG 311

45 Query: 301 QVLAKGPSINPHTRFKGEVYIILSKEEGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEM 360
 QV+AKP SINPHT+FKGEVYILSK+EGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEM
 Sbjct: 312 QVIAKPSSINPHTKFKGEVYIILSKDEGGRHTPFFNNYRPQFYFRTTDVTGSIELPAGTEM 371

Query: 361 VMPGDNVNTIEVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 404
 VMPGDNVTI VELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA
 Sbjct: 372 VMPGDNVNTINVELIHPIAVEQGTTFSIREGGRTVGSGIVSEIEA 415

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 317

A DNA sequence (GBSx0346) was identified in *S.agalactiae* <SEQ ID 1023> which encodes the amino acid sequence <SEQ ID 1024>. Analysis of this protein sequence reveals the following:

55 Possible site: 36
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -0.64 Transmembrane 90 - 106 (90 - 106)

----- Final Results -----

60 bacterial membrane --- Certainty=0.1256 (Affirmative) < succ>

-402-

```
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 318

A DNA sequence (GBSx0347) was identified in *S.agalactiae* <SEQ ID 1025> which encodes the amino acid sequence <SEQ ID 1026>. This protein is predicted to be ftsW. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
15      INTEGRAL    Likelihood = -11.15    Transmembrane   44 - 60 ( 35 - 70)
      INTEGRAL    Likelihood = -4.73    Transmembrane   76 - 92 ( 74 - 98)
      INTEGRAL    Likelihood = -3.88    Transmembrane  117 - 133 ( 113 - 134)

----- Final Results -----
20      bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB39929 GB:U58049 putative cell division protein ftsW
25      [Enterococcus hirae]
      Identities = 78/159 (49%), Positives = 107/159 (67%), Gaps = 4/159 (2%)

Query: 1 MANSXYAMSNGGWFGRGLGNSIEKLGYLPEATTDFVFSIVIEELGVIGAGFILALVFFLI 60
        M+NS YA+ NGG FGRG+GNSI K GYLPE+ TDF+FS++ EE G+IGA +L L+F L
30      Sbjct: 240 MSNSYYALYNGGLFGRGMGNSTITKKGYLPESETDFIFSVIAEEFGLIGALLVLFLFLLC 299

Query: 61 LRIMHVGIKAKDPFNMSIALGIGAMLLMQVFVNIGGISGLIPSTGVTFPFLSQGGNSLLV 120
        +RI K K+ ++I +G+G +L+Q +NIG I GLIP TGV PF+S GG S L+
35      Sbjct: 300 MRIFQKSTKQKNQQANLILIGVGTWILVQTSINIGSILGLIPMTGVPLPFVSYGGTSYLI 359

Query: 121 LSVAIGFVLNIDANEKKELIMKEAEEQYKPQEKENKEIIN 159
        LS AIG LNI + + KE + ++ + Q K K++N
      Sbjct: 360 LSFAIGLALNISSRQVKE---KNKQVERLQLKKPKLLN 394
```

- 40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1027> which encodes the amino acid sequence <SEQ ID 1028>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
45      INTEGRAL    Likelihood = -10.93    Transmembrane  312 - 328 ( 303 - 338)
      INTEGRAL    Likelihood = -8.23    Transmembrane  22 - 38 ( 17 - 47)
      INTEGRAL    Likelihood = -6.85    Transmembrane 192 - 208 ( 187 - 211)
      INTEGRAL    Likelihood = -5.10    Transmembrane 218 - 234 ( 212 - 236)
      INTEGRAL    Likelihood = -4.83    Transmembrane  86 - 102 ( 85 - 107)
      INTEGRAL    Likelihood = -3.72    Transmembrane 385 - 401 ( 383 - 402)
50      INTEGRAL    Likelihood = -3.45    Transmembrane  61 - 77 ( 61 - 79)
      INTEGRAL    Likelihood = -2.39    Transmembrane 344 - 360 ( 344 - 360)

----- Final Results -----
55      bacterial membrane --- Certainty=0.5373 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

-403-

The protein has homology with the following sequences in the databases:

```
>GP:CAB59721 GB:AJ250603 FtsW protein [Enterococcus faecium]
  Identities = 131/397 (32%), Positives = 223/397 (55%), Gaps = 23/397 (5%)
5   Query: 15 KRHLLNYSILLPYLILSVIGLIMVYSTTSVSLIQAHANPKFSVINQGVFWIISLVAITFI 74
        KR +++ IL PYL LS+IGL+ VVS +S L+QA N ++ Q +F +S I
      Sbjct: 3 KRKKIDWWILGPYLTLSMIGLLEVSYASSYRLLQADENTKSLLLRLQLIFIFLSWGVIFLA 62

10  Query: 75 YKLKLNFLNTNTRVLTVVMLGEAFLIIAR--FFTAAIKGAHGWIVIGPVSFQPAEYLKII 132
        +KL++L + ++ + F LI+ R F + GA WI + + FQP+E +
      Sbjct: 63 RSIKLHYLLHPKIAGYGLALSIFFLILVRVGIFGVTVNGAQRWISLFGIQFQPSELANLF 122

15  Query: 133 MVWYLLALTFAKIQKNISLYDYQALTRRKWWPTQWNLDLRDWRVYSLLMVLLVAAQPDLGNA 192
        +++YL+ F P + +L+ + ++ + LL+ QP + A
      Sbjct: 123 LIFYLSWFFRDGN-----PPK--NLKKPFLITVSITLLLILFQPQPKIAGA 164

20  Query: 193 SIIVLTAAIMFSISGIGYRWFSAILVMITGLSTVFLGTIAVIGVERVAKIP-VFGYVAKR 251
        +I+ A ++F + + ++ + +V + L G + +G + +P +F + +R
      Sbjct: 165 LMILSIAWVIFWAAAVPFKKGIYLIVTFSALLIGAAGGVLYLGNK--GWLPQMFNHAYER 222

25  Query: 252 FSAFFNPFHDLTDGHLQANSYYAMSNGGWFGQGLGNSIEKRGYLPEAQTDVFVSVVIEE 311
        + +PF D +G+Q+ +S+YA+ NGG +G+GLGNSI K+GYLPE +TDF+FS++ EE
      Sbjct: 223 IATLDRDPFIDSHGAGYQMTHSFYALYNGGIWGRGLGNSITKKGYLPETEDFIFSIIITEE 282

30  Query: 312 LGLIGAGFILALVFFLILRIMNVGIKAKNPFNAMMALGVGGMMLMQVFVNIGGISGLIPS 371
        LGLIGA +L L+F L +RI + + KN + LG G ++ +Q +N+G I+GL+P
      Sbjct: 283 LGLIGALCVLFLLFSLCMRIFCLSSRCKNQQAGLFLLGFTLFLVQTIMNVGSIAGLMPM 342

35  Query: 372 TGVTFPFLSQGGNSLLVLSVAVGFVLNIDASEKRDDI 408
        TGV PF+S GG S L+LS+ +G INI + + ++
      Sbjct: 343 TGVPLPFVSYGGTSYLISLGIGITLNISSSKIQAEEL 379
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 130/166 (78%), Positives = 152/166 (91%), Gaps = 2/166 (1%)
35  Query: 1 MANSXYAMSNGGWFGRGLGNSIEKRGYLPEAQTDVFVSVVIEELGVIGAGFILALVFFLI 60
        +ANS YAMSNGGWFG+GLGNSIEKRGYLPEAQTDVFVSVVIEELGVIGAGFILALVFFLI
      Sbjct: 269 LANSYYAMSNGGWFGQGLGNSIEKRGYLPEAQTDVFVSVVIEELGVIGAGFILALVFFLI 328

40  Query: 61 LRIMHVGIIKAKDPFNSMIALGIGAMLLMQVFVNIGGISGLIPSTGVTFPLSQGGNSLLV 120
        LRIM+VGIIKAK+PFPN+M+ALG+G M+LMQVFVNIGGISGLIPSTGVTFPLSQGGNSLLV
      Sbjct: 329 LRIMNVGIIKAKNPFNAMMALGVGGMMLMQVFVNIGGISGLIPSTGVTFPLSQGGNSLLV 388

45  Query: 121 LSVAIGFVLNIDANEKKELIMKEAEEQYK--PQEKENKIINLDAFK 164
        LSVA+GFVLNIDA+EK++ I KEAE Y+ + +N K++N+ F+
      Sbjct: 389 LSVAVGFWLNIDASEKRDDIFKEAELSYRKDTRKENSKVNIKQFQ 434
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 319

A DNA sequence (GBSx0348) was identified in *S.agalactiae* <SEQ ID 1029> which encodes the amino acid sequence <SEQ ID 1030>. This protein is predicted to be probable cell division protein ftsw (ftsW). Analysis of this protein sequence reveals the following:

```
Possible site: 34
55  >>> Seems to have an uncleavable N-term signal seq
    INTEGRAL Likelihood = -9.77 Transmembrane 12 - 28 ( 7 - 37)
    INTEGRAL Likelihood = -7.22 Transmembrane 76 - 92 ( 74 - 97)
    INTEGRAL Likelihood = -6.53 Transmembrane 182 - 198 ( 178 - 201)
    INTEGRAL Likelihood = -4.62 Transmembrane 51 - 67 ( 46 - 69)
60  INTEGRAL Likelihood = -2.87 Transmembrane 202 - 218 ( 202 - 218)
```

-404-

----- Final Results -----

bacterial membrane --- Certainty=0.4906 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 5 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9327> which encodes amino acid sequence <SEQ ID 9328> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAA44490 GB:X62621 ORF2 N-terminal [Lactococcus lactis]
 Identities = 82/199 (41%), Positives = 122/199 (61%), Gaps = 9/199 (4%)
 Query: 1 MKIDKRHLLNYSILIPYLILSILGLIVIYSTTSATLIQLGANPFRSVINQGVFWAVSLVA 60
 M ++K + LNYSILIPYLIL+ +G++I+STT +Q G NP++ VINQ F +S++
 15 Sbjct: 1 MNLNKNNFLNYSILIPYLILAGIGIVMIFSTTVPDQLQKGLNPYKLVINQTAFVLLSIIM 60
 Query: 61 IIFIYKLKLNLKNSKVLTMAVLVEVFLLLIARF-----FTQEVNNGAHGWIVIGPI-SF 113
 I IY+LKL LKN K++ + ++ + L+ R T VNGA GWI I I +
 Sbjct: 61 IAVIYRLKLRLALKNRKMIGIIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTV 120
 20 Query: 114 QPAEYKLVIIWYLAFTFARROKKIEIYDYQALTKGRWLPRSLSDLKDWRFYSLFMIGLV 173
 QPAE+ KV I+WYLA F+ +Q++IE D + KG+ L + L WR + ++ +
 Sbjct: 121 QPAEFAKVFIIWYLASFSTKQEEIEKNDINEIFKGKTLTQKL--FGGWRLPVVAILLVD 178
 25 Query: 174 IAQPDILGNGSIIIVLTVIIM 192
 + PDLGN II +IM
 Sbjct: 179 LIMPDLGNTMIIGAVALIM 197

There is also homology to SEQ ID 1028.

30 A related GBS gene <SEQ ID 8545> and protein <SEQ ID 8546> were also identified. Analysis of this protein sequence reveals the following:

Lipop: Possible site: -1 Crend: 6
 McG: Discrim Score: 15.18
 GvH: Signal Score (-7.5): -3.58
 35 Possible site: 34
 >>> Seems to have an uncleavable N-term signal seq
 ALOM program count: 5 value: -9.77 threshold: 0.0
 INTEGRAL Likelihood = -9.77 Transmembrane 12 - 28 (7 - 37)
 INTEGRAL Likelihood = -7.22 Transmembrane 76 - 92 (74 - 97)
 40 INTEGRAL Likelihood = -6.69 Transmembrane 210 - 226 (201 - 227)
 INTEGRAL Likelihood = -6.53 Transmembrane 182 - 198 (178 - 201)
 INTEGRAL Likelihood = -4.62 Transmembrane 51 - 67 (46 - 69)
 PERIPHERAL Likelihood = 1.32 116
 modified ALOM score: 2.45
 45 *** Reasoning Step: 3

----- Final Results -----

bacterial membrane --- Certainty=0.4906 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

ORF02700 (301 - 876 of 1377)
 55 EGAD|8615|8419(1 - 197 of 198) hypothetical protein in rmpg 3'region , fragment {Lactococcus lactis} SP|P27174|YRG2_LACLA HYPOTHETICAL PROTEIN IN RPMG 3'REGION (ORF2) (FRAGMENT). GP|44069|emb|CAA44490.1||X62621 ORF2 N-terminal {Lactococcus lactis} PIR|PC1134|PC1134 hypothetical protein 198 (rmpG 3' region) - Lactococcus lactis (fragment)
 %Match = 15.1
 60 %Identity = 42.3 %Similarity = 64.9
 Matches = 82 Mismatches = 64 Conservative Sub.s = 44

	87	117	147	177	207	237	267	297
	KA*I*I**L*LVILEFLPFFINFL*IYLTGLND*NVPSNISN*SFIFVISIVGGYXX*LIXXXIMHNGNFKY*RK*Y							
5	327	357	387	417	447	477	507	537
	NMKIDKRHLLNYSILIPYLILSILGLIVIYSTSATLIQLGANPFRSVINQGVFWAVSLVAIIIFTYKLKLNFLKNSKVLT :: : : : : ::: : : :: : :: : : : MNLNKNNFLNYSILIPYLILAGIGIVMIFSTTVPDQLQKGGLNPYKLVINQTAFLVLLSIIMIAVITYRLKLRAKKNRKMIG							
10	10	20	30	40	50	60	70	
	567	585	609	636	666	696	726	756
	MAVLVEVFLLLIARF----FT--QEVNAGAHGWIVIGPI-SFQPAEYLKVIIWWYLAFTFARRQKKIEIYDYQALTKGRL : :: : : : : : : : : : : IIIMVILILSLIFCRIMPSSFALTAPVNGARGWIHIPGIGTVQPAEFAKVFIIWYLASFVSTKQEEIEKDNEIFKGKTL							
15	90	100	110	120	130	140	150	
	786	816	846	876	906	936	966	996
	PRSLSDLKDWRFYSLFMIGLVIQPDGLNGSIIIVLTVIIMYCISGIGYRWFSAALLGLIVVGSTLFIGTIAVVGVTMAKV : : : : : : : : : : TQKL--FGGWRLPVVAILLVDLIMPDGLGNITMIIGAVALIMI							
20	170	180	190					

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 320

25 A DNA sequence (GBSx0349) was identified in *S.agalactiae* <SEQ ID 1031> which encodes the amino acid sequence <SEQ ID 1032>. Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence

30 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.3665 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1033> which encodes the amino acid sequence <SEQ ID 1034>. Analysis of this protein sequence reveals the following:

Possible site: 54
>>> Seems to have no N-terminal signal sequence

40

results ----
bacterial cytoplasm --- Certainty=0.2373 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

45 A 1-dimensional 64 kb GAS = 1 GBS with 1 mm⁻¹ 1-dimension

Journal of Oral Rehabilitation 2003; 30: 1023–1030 © 2003 Blackwell Publishing Ltd

50 Query: 1 MEKEAKQIIDLKRNLFKIDVRAQKDEEKVFMRATTCCYSPFY 41
+EKEAKQ+IDLKRNLFKIDVRAQKDEEKVFMRATC S Y
Sbjct: 1 LEKEAKOMIDLKRNLFKIDVRAQKDEEKVFMRATACROSRY 41

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 321

A DNA sequence (GBSx0351) was identified in *S.agalactiae* <SEQ ID 1037> which encodes the amino acid sequence <SEQ ID 1038>. Analysis of this protein sequence reveals the following:

```

5 Possible site: 49
    >>> Seems to have no N-terminal signal sequence
    INTEGRAL Likelihood = -1.65 Transmembrane 78 - 94 ( 78 - 95)
    INTEGRAL Likelihood = -1.33 Transmembrane 421 - 437 ( 420 - 437)

10 ----- Final Results -----
    bacterial membrane --- Certainty=0.1659 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```

15 >GP:CAA00827 GB:A09073 phosphoenol pyruvate carboxylase
      [Corynebacterium glutamicum]
      Identities = 335/958 (34%), Positives = 539/958 (55%), Gaps = 80/958 (8%)

20 Query: 22 EIIIEEVGLLKQLLDEATQKLIGSESFDKIE--KIVSLSLTD---DYTEGLKETISALSNE 76
      + + +++ L Q+L E + G E ++ +E ++ S + + L + ++
      Sbjct: 3 DFLRDDIRFLGQIILGEVIAEQEGQEVYELVEQARLTSFDIAKGNAEMDSLQVFDGITPA 62

25 Query: 77 EMVIVSRYFSILPLLINISEDVDLAYEINYKNNLNQDYLGKLST---TIDVV----- 125
      + + +R FS IL N++ED+ Y L + L T T+D
      Sbjct: 63 KATPIARAFSHFALLANLAEDL-----YDEELREQALDAGDTPPDSTLDATWLKLNEG 115

30 Query: 126 -AGHENAKDILEHVNVVPVLTAHPTQVQRKTVELTSKIHDLRKYRDVKAGIVNQ--- 180
      G E D+L + V PVLTAHPT+ +R+TV + I +R+ ++
      Sbjct: 116 NVGAEAAVADVLRNAEVAPVLTAHPTETRRRTVFDQKWITTHMRERHALQSAEPTARTQS 175

35 Query: 181 --EKWYADLRRYIGIIMQTDITIREKKLVKVNKEITNVMEYYNRSLIKAVTKLTAEYKALAA 238
      ++ ++RR I I+ QT IR + ++++EI + YY SL++ + ++ +
      Sbjct: 176 KLDEIEKNIIRRITILWQTALIRVARPRIEDIEVGLRYYYKLSLLEEIPRINRDVAVELR 235

40 Query: 239 KK---GITHLENPKPLTM-GMWIGGDRDGPNPFTAETLRLSAMVQSEVIINHYTEQINELY 294
      ++ G+ L KB+ G WIGGD DGNP+VTAET+ S +E ++ +Y QL+ L
      Sbjct: 236 ERFGEGVPL---KPVVKPGSWIGGDHDGNPYVTAETVEYSTHRAAETVLKYYARQLHSLE 292

45 Query: 295 RNMMSLISINLTEVSPELVTLANQSDNSVYRENEPYRKAFNFQIQQDKLVQQTLLNLKVGSSPK 354
      +SLS + +V+P+L+ LA+ ++ R +EPYR+A + ++ +++ T
      Sbjct: 293 HELSLSDRMNKVTPQLLALADAGHNDVPSRVDEPYRRAVHGVRGRILAT----- 341

50 Query: 355 EKFVSRQESSDIVGRYIKSHIAQVASDIQTEELPAYATAEEFKQDLLLKVQSLVQYGQDS 414
      ++++++G + + YA+ EEF D L + SL +
      Sbjct: 342 -----TAELIGE-----DAVEGVWFKVFTPYASPEEFLNDAILTIDHSLRESKDVL 386

55 Query: 415 LVDGELACLIQAVDIFGFYLATIDMRQDSSINEACVAELLKSANIVDDYSSLSEEKQL 474
      + D L+ LI A++ FGF L +D+RQ+S E + EL + A + +Y LSE EK ++
      Sbjct: 387 IADDRLSVLISAIESTFGFNLYALDLRQNSESVDLTELFERAQVTANYRELSEAEKLEV 446

500 Query: 475 LLKELTEDPRTLSSTHAPKSELLQKELAIFQTAERELKDQLGEDIIINQHIIISHTESVSDMF 534
      LLKEL + SE+ +EL IF+TA E + G ++ IIS SV+D+
      Sbjct: 447 LLKELRSPRPLIPHGSDEYSEVTDRELGIFRTASEAVKKFGPBMVPHCIISMASVTDVL 506

550 Query: 535 ELAIMLKEVGLIDAN---QARIQIVPLFETIEDLDNSRDIMTQYLHYELVKKWIATNNN 590
      E ++LKE GLI AN + + ++PLFETIEDL I+ + +L + ++ +N
      Sbjct: 507 EPMVLLKEFGLIAANGDNPRGTVDIPLFETIEDLQAGAGILDELWKIDLYRNYYLLQRDN 566

600 Query: 591 YQEIMLGYSDSNKGGYLSSGWTLYKAQNELTKIGEENGIKITFFHGRGGTVGRGGGPSY 650
      QE+MLGYSDSNKGGY S+ W LY A+ +L ++ G+K+ FHGRGGTVGRGGGPSY
      Sbjct: 567 VQEVMLGYSDSNKGGYFSANWALYDAELQVELCRSAGVKLRLFHGRGGTVGRGGGPSY 626

650 Query: 651 EAITSQPFGSTKDRIRLTEQGEIIENKYGNQDAAYNLEMLISASIDRMVTRMITNPNEI 710
      +AI +QP G+++ +R+TEQGEII KYGN + A NLE L+SA+++ + + +E+

```

-407-

Sbjct: 627 DAILAQPRGAVQGSVRITEQGEIISAKYGNPETARRNLEALVSATLE---ASLLDVSEL 682

Query: 711 DNFRETMDGIVSESNAY---YRNLVFDNPYFYDYFFEASPIKEVSSLNIGSRPAARKTI 766
+ + D I+SE + + Y +LV ++ F DYF +++P++E+ SLNIGSRP++RK

5 Sbjct: 683 TDHQRAYD-IMSEISESLKKYASLVHEDQGFIDYFTQSTPLQEIGSLNIGSRPSSRKQT 741

Query: 767 TEISGLRAIPWVFSWSQNRIMPFPGWYGVGSAFKHFI---EQDEANLAKLQTMYQKWPFFN 823
+ + LRAIPWV SWSQ+R+M PGW+GVG+A + +I EQ +A+LQT+ + WPFF

10 Sbjct: 742 SSVEDLRAIPWVLSWSQSRVMLPGWFGVGTALEQWIGEGERQATORIAELQTLNESWPFFT 801

Query: 824 SLLSNVDMLVLSKSNNMIALQYAQLAGSKEVRD-VFNIIILNEWQLTKDMILAIEQHDNLLE 882
S+L N+ V+SK+ + +A YA L EV + V+++I E+ LTK M I D+LL+

Sbjct: 802 SVLDNMAQVMSKAELRIAKLYADLIPDTVEAERVYSVIREEYFLTKKMFCVITGSDDLLD 861

15 Query: 883 ENPMLHASLDYRLPYFNVLNYYQIELIKRLRSNQLDEDYEVKLIHITINGIATGLRNSS 940
+NP+L S+ R PY LN +Q+E+++R R E + I +T+NG++T LRNSG

Sbjct: 862 DNPLLARSVQRYPYLLPLNVIQVEMMRYYRKGDQSEQVSRNIQLTMNGLSTALRNSS 919

A related GBS nucleic acid sequence <SEQ ID 10961> which encodes amino acid sequence <SEQ ID 20 10962> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1039> which encodes the amino acid sequence <SEQ ID 1040>. Analysis of this protein sequence reveals the following:

Possible site: 40

>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1613 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

30 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 659/927 (71%), Positives = 779/927 (83%), Gaps = 11/927 (1%)

35 Query: 14 KLESSSNKEIITEEVGLLKQQLDEATQKLIGSESFDKIEKIVSLSLTDDYTGLKETISAL 73
KLESS+N++II EEV LLK++L+ T+++IG ++F IE I+ LS DY L++ ++ +
Sbjct: 5 KLESSNNQDIIAEDEVALLKEMLENITRRMIGDDAFTVIESIMVLSEKQDYIELEKVVANI 64

Query: 74 SNEEMVIVSRYFSILPLLINISEDVDLAYERINYKNNLNQDYLGKLSTTIDVVAGHENAKD 133
SN+EM ++SRYFSILPLLINISEDVDLAYERINY>NN + DYLGKL+ TI +AG +N KD
40 Sbjct: 65 SNQEMEVISRYFSILPLLINISEDVDLAYERINYQNNTDTDYLGKLALTICKLAGKDNGKD 124

Query: 134 ILEHVNVVPVLTAHPTQVQRKTVELETSKIHDLRLKYRDVKAGIVNQEKWYADLRRYIGI 193
ILE VNVVPVLTAHPTQVQRKT+LELT+ IH LLRKYRD KAG++N EKW +L RYI +
Sbjct: 125 ILEQVNVPVLTAHPTQVQRKTILELTTHIHKLLRKYRDAKAGVINLEKWRQELYRIEM 184

45 Query: 194 IMQTDTIREKKLKVKNEITNVMEYYNRSLIKAVTKLTAEYKALAACKGITHLENPKPLTMG 253
IMQTD TIREKKL+VKNEI NVM+YY+ SLI+AVTKLT EYK LA K G+ L+NPKP+TMG
Sbjct: 185 IMQTDIIREKKLQVKNEIKNMVQYYDGSLIQAVTKLTTEYKVLQKHGLELDNPKPITMG 244

50 Query: 254 MWIGGDRDGNPFTAETLRLSAMVQSEVIINHYIEQLNELYRNMSLSINLTEVSPVELVTL 313
MWIGGDRDGNPFTAETL LSA VQSEVI+N+YI++L LYR SLS L + + E+ L
Sbjct: 245 MWIGGDRDGNPFTAETLCLSATVQSEVIINYYIDEALYRTFSLSSTLVQPNSEVERL 304

55 Query: 314 ANQSQDNHSVYRENEPYRKAFNFQDKLVQTLLNLKVGSSPKEFKVSRQESSDIVGRYIKS 373
A+ SQD S+YR NEPYR+AF++IQ +L QT + L + + SS + S
Sbjct: 305 ASLSQDQSIYRGNEPYRRRAFYIQSRLKQTQIQLT----NQPAASMSSSVGLNTSAWS 358

60 Query: 374 HIAQVASDIQTEELPAYATAEEFKQDLLLLVKQSLVQYQGDSLVDGELACLIQAVDIFGFY 433
A + + I AY + +FK DL ++QSL+ G +L++G+L ++QAVDIFGF+
Sbjct: 359 SPASLENPIL----AYDSPVDFKADLKAIEQSLLDNGNSALIEGDLREVMQAVDIFGFF 413

Query: 434 LATIDMRQDSSINEACVAELLK SANIVDDYSSLSEEKCQCOLLLKELTEDPRTLSSTHAPK 493
LA+IDMRQDSS+ EACVAELLK ANIVDDYSSLSE EKC +LL++L E+PRTLSS K

5 Sbjct: 414 LASIDMRQDSSVQEACVAELLKGANIVDDYSSLSETEKCDVLLQQLMEEPRTLSSAAVAK 473
 Query: 494 SELLOKELAIFQTARELKDLQLEDIINQHISHTESVSDMFELAIMLKEVGLIDANQARI 553
 S+LL+KELAI+ TARELKD+LGE++I QHIISHTESVSDMFELAIMLKEVGL+D +AR+
 Sbjct: 474 SDLLEKELAIYTARELKDKLGEEVIKQHISHTESVSDMFELAIMLKEVGLVDQQRARV 533
 10 Query: 554 QIVPLFETIEDLDNSRDIMTQYLHYELVKKIATNNNYQEIMLGYSDSNKDGYYLSSGWT 613
 QIVPLFETIEDLDN+RDIM YL +++VK WIATN NYQEIMLGYSDSNKDGYYL+SGWT
 Sbjct: 534 QIVPLFETIEDLDNARDIMAAYLSHDIVKSWIATNRNYQEIMLGYSDSNKDGYYLASGWT 593
 Query: 614 LYKAQNELTKIGEENGIKITFFHGRGGTVGRGGGSPSYEAITSQPFGSIKDRIRLTEQGEI 673
 LYKAQNELT IGE+E+G+KITFFHGRGGTVGRGGGSPSY+AITSQPFGSIKDRIRLTEQGEI
 Sbjct: 594 LYKAQNELTAIGEEHGVKITFFHGRGGTVGRGGGSPSYDAITSQPFGSIKDRIRLTEQGEI 653
 15 Query: 674 IENKYGNQDAAYNLEMLISASIDRMVTRMITNPNEIDNFRETMGIVSESNAVYRNLVF 733
 IENKYGN+D AYY+LEMLISASI+RMVT+MIT+PNEID+FRE MD IV++SN +YR LVF
 Sbjct: 654 IENKYGNKDVAYYHLEMLISASINRMVTQMITDPNEIDSFREIMDSIVADSIIYRKLVF 713
 20 Query: 734 DNPFYDYFFEASPIKEVSSLNIGSRPAARKTITEISGLRAIPWVFSWSQRIMFPGWYG 793
 DNP+FYDYFFEASPIKEVSSLNIGSRPAARKTITEI+GLRAIPWVFSWSQRIMFPGWYG
 Sbjct: 714 DNPFYDYFFEASPIKEVSSLNIGSRPAARKTITEITGLRAIPWVFSWSQRIMFPGWYG 773
 Query: 794 VGSAFKHFIEQDEANLAKLOTMYQKWPFFNSLLSNVDMVLSKSNMNLQYAQLAGSKEV 853
 VGSAFK +I++ + NL +LQ MYQ WPFF+SLLSNVDMVLSKSNMNLQYAQLA +V
 25 Sbjct: 774 VGSAFKRYIDRAQGNLERLQHMYQTWPFFHSLLSNVDMVLSKSNMNLQYAQLAERQDV 833
 Query: 854 RDVFNIILNEWQLTKDMILAIQHDNLLEENPMLHASLDYRLPYFNVLNYQIELIKRLR 913
 RDVF IL+EWQLTK++ILAI+ HD+LLE+NP L SL RLPYFNVLNY+QIELIKR R
 Sbjct: 834 RDVFYEILDEWQLTKNVILAIQDHDDILLEDNPSLKHSLKSRLPYFNVLNYIQIELIKRWR 893
 30 Query: 914 SNQLDDEDYEKLIHITINGIATGLRNSG 940
 +NQLE+ EKLIH TINGIATGLRNSG
 Sbjct: 894 NNQLDENDEKLIHTTINGIATGLRNSG 920

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 322

A DNA sequence (GBSx0352) was identified in *S.agalactiae* <SEQ ID 1041> which encodes the amino acid sequence <SEQ ID 1042>. This protein is predicted to be Bacillus licheniformis Pz-peptidase homologue (pepF). Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
45 bacterial cytoplasm --- Certainty=0.3012 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1043> which encodes the amino acid sequence <SEQ ID 1044>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
55 bacterial cytoplasm --- Certainty=0.3137 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

-409-

Identities = 512/593 (86%), Positives = 564/593 (94%)

Query: 1 MKLKKRSEFPENELWDLITALYKDRQDFLLAIEKALEDIKVFKNYEGKLNCVEDFTSALM 60
M+LKKRSEFPENELWDLITALYKDRQDFLLAIEKAL+DI +FK+NYEG+L V+DFT AL+

5 Sbjct: 26 MELKKRSEFPENELWDLITALYKDRQDFLLAIEKALQDIDI,FKRNYEGRITSVDDFTQALI 85

Query: 61 EIEHIYIQMSHIDTYAFMPQTTSNEEFQAQISQAGSDFATKANVLLSFFNTALANADIK 120
EIEHIYIQMSHI TYAFMPQTTSNEEFQAQISQAGSDFATKANVLLSFFNTALANAD+

10 Sbjct: 86 EIEHIYIQMSHIGTYAFMPQTTSDESAFAQIAQAGDDFTKASVALSFFDTALANADLD 145

Query: 121 ILDSLENNPHFKATIRQAKIQKQHLLSPEVEKALTNLNEVNTPYDIYTCKMRAGDFDMED 180
+LD+LE NP+F A IR AKIQK+HLLSP+VEKAL NL EV+N PYDIYTCKMRAGDFDM+D

Sbjct: 146 VLDTLEKNPYFSAAIRMAKIQKEHLLSPDVEKALANLREVINAPYDIYTCKMRAGDFDMDD 205

15 Query: 181 FEVDGKTYKNSFVTYENYFQNHENAEIREKSFRSFSKGRLKHQNAAAAYLAKVKSEKLI 240
FEVDGKTYKNSFV+YEN++QNHENAEIREK+FRSFSKGRLKHQN AAAAYLAKVKSEKLI+

Sbjct: 206 FEVDGKTYKNSFVSYENFYQNHENAEIREKAFRSFSKGRLKHQNAAAAYLAKVKSEKLL 265

20 Query: 241 ADMRGYDSVFDYLSEQEVDRSMFDRQIDLIMDEFGPVAQRFLKHIADVNGIEKMTFADW 300
ADM+GY SVFDYLL+EQEVDRS+FDRQIDLIM EFGPVAQ+FLKH+A VNG+EKMTFADW

Sbjct: 266 ADMKGYASVFDYLLAEQEVDRSLSFDRQIDLIMTEFGPVAQKFLKHVAQVNGLERKMTFADW 325

Query: 301 KLDIDNELNPEVSINDAYDLVMKSVAPLGLKEYSQEVERYQKERWVDAANANKDSGGYAA 360
KLDIDDN+LNPEVSI+ AYDLVMKSL+APLG+EY++E+ERYQ ERWVDAANANKDSGGYAA

25 Sbjct: 326 KLDIDNDLNPEVSIDGAYDLVMKSLAPLQETYKIEIERYQTERWVDAANANKDSGGYAA 385

Query: 361 DPYKVHPYVLMSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSFFNTHMSTYYVEAPSTFNE 420
DPYKVHPYVLMSWTGRMSDVYTLIHEIGHSGQFIFSDNHQS+FNTHMSTYYVEAPSTFNE

30 Sbjct: 386 DPYKVHPYVLMSWTGRMSDVYTLIHEIGHSGQFIFSDNHQSFYFNTHMSTYYVEAPSTFNE 445

Query: 421 LLLSDYLENQFDTARQKRFALAHRLTDYFHNFITHLLEAAFQRKVYTLIEEGGTFGAEQ 480
L+LSDYLE+QFD RQKRFALAHRLTDYFHNFITHLLEAAFQRKVYTLIEEGGTFGA+Q

Sbjct: 446 LMLSDYLEHQFDDPRQKRFALAHRLTDYFHNFITHLLEAAFQRKVYTLIEEGGTFGAQDQ 505

35 Query: 481 LNAIMKEVLTOFWGDAIEIDDAALTWMRQAHYYMGLYSYTYTAGLVISTAGYLNLKNNP 540
LNA+MKEVLT FWGDA++IDDAALTWMRQAHYYMGLYSYTYTAGLVISTAGYLNLK+NP

Sbjct: 506 LNAMMKEVLTDFWGDAVIDDDAALTWMRQAHYYMGLYSYTYTAGLVISTAGYLNLKHNP 565

40 Query: 541 NGAKEWLAFLKSGGSRTPLETALLISADISTDKPLRDTINFLSNTVDQIINYS 593

NGAKEWL FLKSGGSRTPL+TA+LI ADI+T+KPLRDTI FLS+TVDQII+Y+

Sbjct: 566 NGAKEWLDFLKSGGSRTPLDTAMLIGADIATEKPLRDTIQFLSDTVDQIISYT 618

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

45 Example 323

A DNA sequence (GBSx0353) was identified in *S.agalactiae* <SEQ ID 1045> which encodes the amino acid sequence <SEQ ID 1046>. Analysis of this protein sequence reveals the following:

Possible site: 19

>>> May be a lipoprotein

50 ----- Final Results -----

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

55

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1047> which encodes the amino acid sequence <SEQ ID 1048>. Analysis of this protein sequence reveals the following:

60 Possible site: 19

-410-

>>> May be a lipoprotein

----- Final Results -----

5 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 72/127 (56%), Positives = 85/127 (66%)

Query: 1 MKKYIKLFLLTVFATTLVACGQPSTSINKTTSSSTLEVGVKELVVKEDETNVLSEKVYHKG 60
 + K K L + A LVAC Q + +TT S V LVVKEDETN + EKV + KG

Sbjct: 1 VNKRKFTGFLALVAMLLVACSGTKQIQTTPSVPKADHHVRLVVKEDEKVSFGKG 60

15 Query: 61 DTVLDVDLKANYKVKEKDGFITSIDGISQDETKGlyWMFKVNNKLAPKAANQIKVKNDKI 120
 DTVL+VLK NY+VKEKDGFIT+IDGI QD YW+FKVN K+A K A+QI VK D I
 Sbjct: 61 DTVLEVLKDNYEVKEKDGFITAIDGIEQDTKANKYWLFKVNGKMADKGADQITVKDGDSI 120

20 Query: 121 EFYQEVY 127

EFYQEV+

Sbjct: 121 EFYQEVF 127

SEQ ID 1046 (GBS185) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 25 extract is shown in Figure 28 (lane 6; MW 15.7kDa).

GBS185-His was purified as shown in Figure 199, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 324

30 A DNA sequence (GBSx0354) was identified in *S.agalactiae* <SEQ ID 1049> which encodes the amino acid sequence <SEQ ID 1050>. Analysis of this protein sequence reveals the following:

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

35 INTEGRAL Likelihood = -4.46 Transmembrane 75 - 91 (67 - 94)
 INTEGRAL Likelihood = -4.41 Transmembrane 33 - 49 (30 - 49)
 INTEGRAL Likelihood = -2.60 Transmembrane 53 - 69 (52 - 70)
 INTEGRAL Likelihood = -1.38 Transmembrane 108 - 124 (106 - 124)
 INTEGRAL Likelihood = -0.06 Transmembrane 149 - 165 (149 - 165)

40 ----- Final Results -----

bacterial membrane --- Certainty=0.2784 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45 A related GBS nucleic acid sequence <SEQ ID 9731> which encodes amino acid sequence <SEQ ID 9732> was also identified. A further related GBS nucleic acid sequence <SEQ ID 10929> which encodes amino acid sequence <SEQ ID 10930> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1051> which encodes the amino acid
 50 sequence <SEQ ID 1052>. Analysis of this protein sequence reveals the following:

Possible site: 48

>>> Seems to have a cleavable N-term signal seq.

-411-

INTEGRAL	Likelihood = -7.96	Transmembrane	50 - 66	(49 - 71)
INTEGRAL	Likelihood = -5.73	Transmembrane	101 - 117	(99 - 124)
INTEGRAL	Likelihood = -4.41	Transmembrane	141 - 157	(139 - 159)
INTEGRAL	Likelihood = -4.25	Transmembrane	73 - 89	(67 - 92)

5

----- Final Results -----

bacterial membrane	--- Certainty=0.4185(Affirmative) < succ>
bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

10

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/163 (50%), Positives = 120/163 (73%), Gaps = 3/163 (1%)

15

Query: 10	LTRVAILSALCVVLRYAFAPLPNIQPIATAIFLITVVVLFDLKEGVATVTITMLVSSFLMGF 69
	++R+AI+SALCVVLR F+ LPN+QP+TA L ++ F L E V + + + +S+FL+GF
Sbjct: 6	MSRIAIMSALCVVLRMVFSSLNPNVQPVTAFLLSYLLYFGIAEAVLVMMCLFLSAFLLG 65

20

Query: 70	GWPWFLQIISFTLILCLWKFLIYPLTKAVCFGKITEEVVLQTFFAGGLGVVYGVIIDTCFA 129
	GPWVF Q+ F L+L LW+F++YPL++ F K ++ Q F G++YGV+IDTCFA
Sbjct: 66	GPWVFQVTCFVLVLLLWRFVLYPLSQQ--FPKY-QLGCQAFLVALCGLLYGVLIDTCFA 122

25

Query: 130	WLYHMPWWTYVLAGLSFNMAHALSTCLFYPLLPILRRFRNEK 172
	+LY MPWW+YVLAG+ FN+AHALST +F+P+++ + RR E+
Sbjct: 123	YLYSMPWWSYVLAGMPFNIAHALSTLVFFFVVMMLFRRRLIGEQ 165

A related GBS gene <SEQ ID 8549> and protein <SEQ ID 8550> were also identified. Analysis of this protein sequence reveals the following:

30

Lipop: Possible site: -1 Crend: 10

McG: Discrim Score: 6.79

GvH: Signal Score (-7.5): ~0.91

Possible site: 28

>>> Seems to have a cleavable N-term signal seq.

ALOM program count: 3 value: -4.46 threshold: 0.0

35

INTEGRAL	Likelihood = -4.46	Transmembrane	35 - 51	(29 - 54)
INTEGRAL	Likelihood = -1.38	Transmembrane	68 - 84	(66 - 84)
INTEGRAL	Likelihood = -0.06	Transmembrane	109 - 125	(109 - 125)

PERIPHERAL	Likelihood = 7.53	88
------------	-------------------	----

40

modified ALOM score: 1.39

*** Reasoning Step: 3

----- Final Results -----

bacterial membrane	--- Certainty=0.2784(Affirmative) < succ>
bacterial outside	--- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm	--- Certainty=0.0000(Not Clear) < succ>

45

The protein has homology with the following sequences in the databases:

50

ORF01220 (421 - 552 of 1002)	
GP 9950155 gb AAG07353.1 AE004814_8 AE004814 (16 - 56 of 69)	hypothetical protein
{Pseudomonas aeruginosa}	
%Match = 3.2	
%Identity = 39.5 %Similarity = 60.5	
Matches = 17 Mismatches = 15 Conservative Sub.s = 9	

55

222	252	282	312	342	372	402	432
STLTKLTRVAILSALCVVLRYAFAPLPNIQPIATAIFLITVVVLFDLKEGVATVTITMLVSSFLMGFGPWVFLQIISFTLIL							
::: MDPELFEEWMMTGLVTVLI							

60

10

462	492	522	552	582	612	642	672
CLWKFLIYPLTKAVCFGKITEEVVLQTFFAGGLGVVYGVIIDTCFAWLYHMPWWTYVLAGLSFNMAHALSTCLFYPLLPI							

```

: |::: | | || ::| ||| |||| | ||
LFMAFIIVWDIAKKS KAGKFGTLIL--FFALGLGV-LGFIIKGLVIGSLEGAGM
      30        40        50        60

```

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 325

A DNA sequence (GBSx0355) was identified in *S.agalactiae* <SEQ ID 1053> which encodes the amino acid sequence <SEQ ID 1054>. This protein is predicted to be endolysin. Analysis of this protein sequence 10 reveals the following:

```

Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

```

```
----- Final Results -----
```

```

15 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```

20 >GP:CAA72266 GB:Y11477 endolysin [Bacteriophage Bastille]
      Identities = 64/210 (30%), Positives = 95/210 (44%), Gaps = 15/210 (7%)
      Query: 66 KPIIDVSGWQLPKEIDYDTLSKNISGVVIRVF GGSKISKTNNAAYTTGIDKSFKTHIKEF 125
              K I+D+S +ID+DT +S + R G ++ +N +D+ +KT +
      Sbjct: 12 KTIVDISHHNA--DIDFD TAKNYVSMFIARTGDGHRYN--SNGELQGVVDRKYKTFVANM 67
      Query: 126 QKRNIPVAVSYALGSSVKEMKEEAQIFYKNAAPYKPTFYWIDVEEETMSNMNKGVQA FR 185
              + R IP Y + S V K+EA+ F+ N T + D E T NM + +Q F
      Sbjct: 68 KARGIPFGNMFNRSGVASAKQEAEFFW-NYGDKD ATVVVCDAEVSTAPNMKEC IQVFI 126
      Query: 186 KELKRLGAKNVGIIYIGTYFMTEQGISVKGFD AVWIPTYGSDGYYEAAPQTELKYDLHQY 245
              LK LGAK VG+YIG + E G D WIP YG+ + DL Q+
      Sbjct: 127 DRLKELGAKKVGLIYIGHHKYQEFGGKD VNCDFTWIPRGNKP A-----ACDLWQW 177
      Query: 246 TSQGYLPGFNQPLDLNQIAVN KDKKKTYEK 275
              T G + G + D+N + +K EK
      Sbjct: 178 TEYGNIA GIGK-CDINVLYGD KPMSSFT EK 206

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1055> which encodes the amino acid 40 sequence <SEQ ID 1056>. Analysis of this protein sequence reveals the following:

```

Possible site: 31
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL Likelihood = -16.98 Transmembrane 8 - 24 ( 3 - 28)

```

```

45 ----- Final Results -----
      bacterial membrane --- Certainty=0.7793(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

50 An alignment of the GAS and GBS proteins is shown below:

```
Identities = 198/278 (71%), Positives = 235/278 (84%)
```

```

Query: 1 MRRRIKPIVVAVFFSLFGLLLIGHLHSTNTLK KELVEAKKTI PSVKASKVPQKSTSSKD 60
      MRR+IKPIVV VFF L ++LIIG + + +KE+ +AK IP ++ K++S+
      Sbjct: 1 MRRKIKPIVVLFVFFILLAMVLII GKRQANHAKQKEVEDAKSHI PIATSNPGKAKTSTSET 60
      Query: 61 KEFVLKPIIIDVSGWQLPKEIDYDTLSKNISGVVIRVF GGSKISKTNNAAYTTGIDKSFKT 120
              ++F+L PI+DVSGWQLP+EIDYDTLS++ISG ++RV+GGS+I+ NNAA+TTGIDKSFKT
```


-414-

PAFACDLWQWTEYGNIAIGIK-CDINVLYGDKPMSSFFTEKEGAKETLVPALNKVVTVYEVGTNLIPEIQDKLAFLGYEAR
 180 190 200 210 220 230 240

SEQ ID 8552 (GBS206) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell
 5 extract is shown in Figure 51 (lane 6; MW 31.7kDa).

GBS206-His was purified as shown in Figure 206, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 326

10 A DNA sequence (GBSx0356) was identified in *S.agalactiae* <SEQ ID 1057> which encodes the amino acid sequence <SEQ ID 1058>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
      INTEGRAL   Likelihood = -1.44   Transmembrane 183 - 199 ( 183 - 200)
15
----- Final Results -----
      bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

20 A related GBS nucleic acid sequence <SEQ ID 9729> which encodes amino acid sequence <SEQ ID 9730> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG20117 GB:AE005090 NADH dehydrogenase/oxidoreductase-like
25        protein; NolA [Halobacterium sp. NRC-1]
      Identities = 38/156 (24%), Positives = 83/156 (52%), Gaps = 13/156 (8%)
      Query: 19 TMEILIAGGSGFLGKQIIKAALTKGHKVAYILSRHEGKGDFKDPRLTYIRGDITEADKIH 78
      +M++L+ GG+GF+G + + +GH V +R + D +T I GD+T + +
30        Sbjct: 8 SMDVLVTGGTGFIGTHLCRELDDRGHDVTAFAREPADAALPAD--VTRIVGDVTVKETVA 65
      Query: 79 LEDRTFDILIDCIGA---IKPNQLD---ELNVKATQKAVALCHKNQIPKLVYISA--- 127
      D +++ + KP+ D +++++ T+ VA + + ++ +SA
35        Sbjct: 66 NAIDGHDAVVNLVALSPLFKPKPSGDSRHLDVHLGGTENVAAASEAGVEYILQLSALDAD 125
      Query: 128 NSGYSAYIKSKRKAEQIIKASGLDYLFRVRPGLMYGE 163
      +G +AY+++K +AE+ +++S L + VRP +++G+
      Sbjct: 126 PTGPTAYLRAKGRAEEAVRSSDLHHTIVRPSVVF 161
```

40 No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8553> and protein <SEQ ID 8554> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1 Crend: 5
McG: Discrim Score: -7.99
45        GvH: Signal Score (-7.5): -6.34
      Possible site: 41
      >>> Seems to have no N-terminal signal sequence
      ALOM program count: 1 value: -1.44 threshold: 0.0
      INTEGRAL   Likelihood = -1.44   Transmembrane 183 - 199 ( 183 - 200)
50        PERIPHERAL Likelihood = 4.29        20
      modified ALOM score: 0.79
      *** Reasoning Step: 3
```

----- Final Results -----

bacterial membrane --- Certainty=0.1574 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5

RGD motif 68-70

The protein has homology with the following sequences in the databases:

50 SEQ ID 8554 (GBS303) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 5; MW 28.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 5; MW 53.2kDa).

The GBS303-GST fusion product was purified (Figure 207, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 275), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 327

A DNA sequence (GBSx0357) was identified in *S.agalactiae* <SEQ ID 1059> which encodes the amino acid sequence <SEQ ID 1060>. Analysis of this protein sequence reveals the following:

5 Possible site: 49
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2850 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

15 >GP: AAC36853 GB:L23802 regulatory protein [Enterococcus faecalis]
 Identities = 61/164 (37%), Positives = 96/164 (58%), Gaps = 13/164 (7%)

20 Query: 1 MSKKNKIKKTLVDQILDKAKIEH-----DSLQLDALQGDLPPNGIQQKQDIFKTLALI 51
 M+KK +KT +++++ K+ + D L +++ L GI+K IFKTL +
 Sbjct: 1 MAKKKTQQKTNAMRMVEQHKVPYKEYFAWSEDHLSAESVAESL--GIEKGRIFKTLVTV 58

25 Query: 52 GDKTGPIIGILPLTEHLSEKKLAKISGNKKVQMI P Q K D L Q K I T G Y I H G A N N P I G I R Q K H N 111
 G+KTGP++ ++P + L KKLAK SGNKKV+M+ KDL+ TGYI G +P G+ K
 Sbjct: 59 GNKTGPVVAVIPGNQELDLKSLAKASGNKKVEMLHLKDLEATTGYIRGGCSPTGM--KKQ 116

Query: 112 YPIFIDTIALEKQELIVSAGEIGRSIRINSEVLADFVNNAKFADI 155
 +P ++ A + +IVSAG+ G I + E + N +FA+I
 Sbjct: 117 FPTYLAEEAQYSAAIVSAGKRGMQIELAPEATLSLTNGQFAEI 160

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1061> which encodes the amino acid sequence <SEQ ID 1062>. Analysis of this protein sequence reveals the following:

30 Possible site: 30
 >>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2651 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

40 Identities = 114/157 (72%), Positives = 139/157 (87%)

Query: 1 MSKKNKIKKTLVDQILDKAKIEHDSLQLDALQGDLPPNGIQQKQDIFKTLALI GDKTGPPIIG 60
 M+KK K+KKT L V +Q I L D K A I H L +L+AL+GD P+ +Q D I +K T L A L G D +T G P +I G
 Sbjct: 1 MAKKT K L K K T L V E Q I L D K A N I A H Q G L K L N A L E G D F P D D L Q P S D I Y K T L A L T G D Q T G P L I G 60

45 Query: 61 I I P L T E H L S E K K L A K I S G N K K V Q M I P Q K D L Q K I T G Y I H G A N N P I G I R Q K H N Y P I F I D T I A 120
 I+PLTEHLSEK+LAK+SGNKKV M+P Q K D L Q K T G Y I H G A N N P +G I R Q K H +Y P I F I D A
 Sbjct: 61 I I P L T E H L S E K Q L A K V S G N K K V S M V P Q K D L Q K T T G Y I H G A N N P V G I R Q K H S Y P I F I D Q T A 120

50 Query: 121 LEKQELIVSAGEIGRSIRINSEVLADFVNNAKFADIKE 157
 LEK ++IVSAGE+GRSI+I+S+ LADFV A FAD+K+
 Sbjct: 121 LEKGQIIIVSAGEVGRSIKISSQALADFVGASFADLKK 157

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

55 Example 328

A DNA sequence (GBSx0358) was identified in *S.agalactiae* <SEQ ID 1063> which encodes the amino acid sequence <SEQ ID 1064>. Analysis of this protein sequence reveals the following:

-417-

Possible site: 28
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.4719 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 8555> which encodes amino acid sequence <SEQ ID 8556>
 10 was also identified. This protein belongs to the glycolysis/gluconeogenesis pathway, and such proteins have been experimentally detected as surface-exposed in *Streptococci*. The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD36444 GB:AE001791 phosphoglycerate mutase [Thermotoga maritima]
 15 Identities = 65/191 (34%), Positives = 93/191 (48%), Gaps = 13/191 (6%)
 Query: 5 MKFYLVRHGKTQWNLEGRFQGANGDSPLLEAAIEEELGQYLSSIHFDAYVSSDLGRAR 64
 MK YL+RHG+T WN +G +QG D PL E E+ +L L + DA+YSS L R+
 Sbjct: 1 MKLYLIRHGETIWNNEKGLWQGVVT-DVPLNERGREQARKLANSLKRV--DAIYSSPLKRSL 57
 20 Query: 65 DTVNLLNDANSCKPEIHYTPQLREWALGTLEGCKIATMQATYPRQMTAFYQNPLQFKHDM 124
 +T + A KEI IRE + G + YP + + P M
 Sbjct: 58 ETAEEI--ARRFEKEIIIVEEDLRECEISLWNGLTVVEAIREYPVEFKKWSSDP---NFGM 112
 25 Query: 125 FGAESLYQTTHRVESFLRSLASK---NYDKVLIVGHGANLTASIRSLGYQY GSLHYKD 180
 G ES+ +RV + + S+ + V+IV H +L A I +LG LH
 Sbjct: 113 EGLESMRNVQNRVVKAIMKIVSQEKLINGSENVVIVSHSLSLRAFICWIILGLPL-YLHRNF 171
 Query: 181 KLDNASLTIIE 191
 KLDNASL+++E
 30 Sbjct: 172 KLDNASLSVVE 182

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1065> which encodes the amino acid sequence <SEQ ID 1066>. Analysis of this protein sequence reveals the following:

35 Possible site: 24
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3628 (Affirmative) < succ>
 40 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 127/205 (61%), Positives = 152/205 (73%)

45 Query: 5 MKFYLVRHGKTQWNLEGRFQGANGDSPLLEAAIEEELGQYLSSIHFDAYVSSDLGRAR 64
 MK Y VRHGKT WNLEGRFQGA GDSPLLEAA +E+ LG+ L+ + FDAVY+SDL RA
 Sbjct: 1 MKLYFVRHGKTWNLEGRFQGAGGDSPLEEAKEDEIHLGKELAKVAFDAVYTSQMLRAM 60
 50 Query: 65 DTVNLLNDANSCKPEIHYTPQLREWALGTLEGCKIATMQATYPRQMTAFYQNPLQFKHDM 124
 T I+ DA +--+T QLREW LG LEG KIATM ATYP+QM AF +N QFK D
 Sbjct: 61 ATAIIILDAFDQQPKLYHTDQLREWRLGKLEGAKIATMAATYPOQMLAFRENLAQFKPDQ 120
 Query: 125 FGAESLYQTTHRVESFLRSLASKNYDKVLIVGHGANLTASIRSLGYQY GSLHYKD KLDN 184
 F AES+YQTT RV ++S K+Y VLIVGHGANLT+IRSLLG++ L K LDN
 55 Sbjct: 121 FEAESIYQTQRVCHLIQSFKDKHYQNVLIVGHGANLTATIRSLLGFEPALLLAKGGLDN 180
 Query: 185 ASLTITETHDFKDFNCLTWNDKSYL 209
 ASLTII+ET D+ ++CL WNDKS+L
 Sbjct: 181 ASLTILETKDYLTYDCLIWNDKSFL 205

SEQ ID 8556 (GBS314) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 4; MW 27.2kDa), in Figure 169 (lane 15-17; MW 41.6kDa) and in Figure 239 (lane 4; MW 41.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 4; MW 52.1kDa).

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 329

A DNA sequence (GBSx0359) was identified in *S.agalactiae* <SEQ ID 1067> which encodes the amino acid sequence <SEQ ID 1068>. Analysis of this protein sequence reveals the following:

```
10 Possible site: 56
     >>> Seems to have no N-terminal signal sequence

     ----- Final Results -----
     bacterial cytoplasm --- Certainty=0.3014(Affirmative) < succ>
15      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
20 >GP:CAB12562 GB:Z99108 similar to hypothetical proteins [Bacillus subtilis]
     Identities = 69/232 (29%), Positives = 108/232 (45%), Gaps = 9/232 (3%)

     Query: 4 SIVFDVDDTIYDQQAPYRIAVEKCFPDFMSAINQAYIRFRHYSIDGFPVMAGEWTTEY 63
             +--+FDVDDTI D QA +A+ F D ++ N +++ + + G+ T +
     Sbjct: 6 TLLFDVDDTILDQAAEALALRLLFEDQNIPLTNDMKAQYKTINQGLWRAFEEGKMTTRDE 65

     25 Query: 64 FRFWRCKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRTLDFLKSKNVPMGII 123
             R L E+GY EA G ++ Y LE L + L + + I+
     Sbjct: 66 VVNTRFSALLKEYGY---EADGALLEQKYRRFLEEGHQLIDGAFDLISNLQQQFDLYIV 121

     30 Query: 124 TNGPTEHQLKKVKLGLYDYVDPKRVIVSQATGFQKPEKEIFNLAAEQF-DMNPSTTLVY 182
             TNG + Q K+++ GL+ + K + VS+ TGFQKP KE FN E+ + TL +
     Sbjct: 122 TNGVSHTQYKRLRDSLGLFPFF--KDIVSEDTGFQKPMKEYFNYVFERIPQFSAEHTLII 179

     35 Query: 183 GDSYNDIMGAFNNGGWHSMWFNHRGRSLKPGIKPVYDVAIDNFEQLFGAVKV 234
             GDS DI G G + W N + P I P Y+ I E+L+ + +
     Sbjct: 180 GDSLTDIKGQLAGLDTCWMNPDMKPNVPEI IPTYEE--IRKLEELYHILNI 229
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1069> which encodes the amino acid sequence <SEQ ID 1070>. Analysis of this protein sequence reveals the following:

```
40 Possible site: 56
     >>> Seems to have no N-terminal signal sequence

     ----- Final Results -----
     bacterial cytoplasm --- Certainty=0.3216(Affirmative) < succ>
45      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
50 Identities = 276/300 (92%), Positives = 292/300 (97%)

     Query: 1 MITSIVFDVDDTIYDQQAPYRIAVEKCFPDFMSAINQAYIRFRHYSIDGFPVMAGEWT 60
             MIT+IVFDVDDTIYDQQAPYRIA+EKCFPDFMS +NQAYIRFRHYS+GFPVMAGEWT
     Sbjct: 1 MITAIVFDVDDTIYDQQAPYRIAMEKCFPDFMSVMNQAYIRFRHYSIDGFPVMAGEWT 60

     55 Query: 61 TEYFRFWRCKETLLEFGYREIDEATGIYFQEIYEHELENITMLDEMRTLDFLKSKNVPM 120
             TEYFRFWRCKETLLEFGYREIDEA G++FQE+YEHELENITMLDEMRTLDFLKSKNVPM
```

5 Sbjct: 61 TEYFFWRCKETLLEFGYREIDEAAGVHFQEYEHLENITMLDEMRTMLDFLKSKNVPM 120
 Query: 121 GIITNGPTEHQLKKVKKLGLYDYVDPKRVIVSQATGFQPKPEKEIFNLAAEQFDMNPSTTL 180
 GIITNGPTEHQLKKV+KLGLYDY+D KRVIVSQATGFQPKPEKEIFNLAAEQFDMNP TTL
 Sbjct: 121 GIITNGPTEHQLKKVRLGLYDYIDAKRVIVSQATGFQPKPEKEIFNLAAEQFDMNPQTTL 180
 Query: 181 YVGDSYDNDIMGAFNGGWHSWFNHRGRSLKPGIKPVYDAIDNFEQLFGAVKVLFDLDP 240
 YVGDSYDNDIMGAFNGGWHSWFNHRGR LKPG KPVYDAIDNFEQLFGAVKVLFDLDP
 Sbjct: 181 YVGDSYDNDIMGAFNGGWHSWFNHRGRQLPGTKPVYDAIDNFEQLFGAVKVLFDLDP 240
 Query: 241 NKFIFDINDKSNPVLEMGLNNGLMMAAERLLESNMSVDKVILLRLTAKQEKVLRMKYAR 300
 NKFIFD+NDK NP+L+MG+NNGLMMAAERLLESNMS+DKVILLRLT +QEKVLR+KYAR
 Sbjct: 241 NKFIFDVNDKKNPILQMGINNGLMMAAERLLESNMSIDKVILLRLTAKQEKVLRMKYAR 300

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 330

A DNA sequence (GBSx0360) was identified in *S.agalactiae* <SEQ ID 1071> which encodes the amino acid sequence <SEQ ID 1072>. Analysis of this protein sequence reveals the following:

20 Possible site: 34
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2451(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

25 A related GBS nucleic acid sequence <SEQ ID 9727> which encodes amino acid sequence <SEQ ID 9728> was also identified.

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB11858 GB:Z99104 lysyl-tRNA synthetase [Bacillus subtilis]
 Identities = 318/490 (64%), Positives = 390/490 (78%), Gaps = 1/490 (0%)
 Query: 44 EELNDQQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDLHDIEETATIA 103
 EELNDQ VRR+KM L + GIDPFG RFERT S ++ Y D +KE+L + TIA
 Sbjct: 9 EELNDQLQVRRDKMNQLRDNGIDPFGARFERTHQSQEVISAYQDLTKEELEKAIEVTIA 68
 Query: 104 GRLMTKRGKGKVGFAHIQDREGQIQTIVRKDSVGEENEYEIFKKADLGDFLGVEQVMRTD 163
 GR+MTKRGKGK GFAH+QD EGQIQTIVRKDSVG++ YEIFK +DLGD +GV G+V +T+
 Sbjct: 69 GRMMTKRGKGKAGFAHLQDLEGQIQTIVRKDSVGDDQYEIFKSSDLGDLIGVTGKVFKTN 128
 Query: 164 MGELSIKATHITHLSKALRPLPEKFHGLTDIETIYRKRHLDLISNRDSFDRFVTRSKIIS 223
 +GELS+KAT L+KALRPLP+K+HGL D+E YR+R+LDL I N DS F+TRSKII
 Sbjct: 129 VGELSVKATSFELLTKALRPLPDFKYHGLKDVQRYRQRYLDLIVNPDSKHTFITRSKIIQ 188
 Query: 224 EIRRFMDNSGFLFLEVEPTVLHNEAGGASARPFITHHNAQIDDMVLRIATELHLKRLIVGGM 283
 +RR++D +G+LEVETP +H+ GGASARPFITHHNA DI + +RIA ELHLKRLIVGG+
 Sbjct: 189 AMRYYLDDHGYLEVETPTMHISIPGGASARPFITHHNALDIPLYMRIAIELHLKRLIVGG 248
 Query: 284 ERVYEIGRIFRNEGMMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVTKTVKGDGPIN 343
 E+VYEIGR+FRNEG+ HNPEFT IE Y+AYADY+DIM LTE ++ H+ + V G I
 Sbjct: 249 EKVYEIGRVFRNEGVSTRHNPEFTMIELYEAYADYKDIMSLENLVAHIAQEVLGTTIQ 308
 Query: 344 YQGTEIKINEPFKRVHMVDAVKEITGIDFWKEMTLEEAQALAQEKNVPLEKHFTTVGHII 403
 Y +I + +KR+HMVDAVKE TG+DFW+E+T+E+A+ A+E V + K TVGHII
 Sbjct: 309 YGEEQIDLKPEWKRRIHMVDAVKEATGVDFWEEVTVQAREYAKEHEVEI-KDSMTVGHII 367
 Query: 404 NAFFEEFVEDTLIQPPTFVFGHPVEVSPLAKKNNDPRTDRFELFIMTKEYANAFTELND 463
 N FFE+ +E+TLIQPPTF++GHPVE+SPLAKKN DPRFTDRFELFI+ +E+ANAFTELND
 Sbjct: 368 NEFFEQQKIEETLQIOPTFIYGHPEVIEPLAKKNPDEDPRFTDRFELFIVGREHANAFTELND 427

-420-

Query: 464 PIDQLSRFEAQASAKELGDDATGVVDYDYEVALEYGMPPPTGGLGIGIDRLCMLLTDTTTI 523
 PIDQ RFEAQ +E G+DEA +D D+VEALEYGMPPPTGGLGIGIDRL MLLT+ +I
 Sbjct: 428 PIDQRERFEAQLKEREAGNDEAHLMDEFVVALEYGMPPPTGGLGIGIDRLVMLLTNPASI 487

5 Query: 524 RDVLLFPTMK 533
 RDVLLFP M+
 Sbjct: 488 RDVLLFPQMR 497

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1073> which encodes the amino acid sequence <SEQ ID 1074>. Analysis of this protein sequence reveals the following:

Possible site: 45
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4694 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

20 An alignment of the GAS and GBS proteins is shown below:

Identities = 439/500 (87%), Positives = 474/500 (94%)

Query: 34 LEEIMSNQHIEELNDQQIVRREKMAALTEQGIDPFGKRFERTATSGQLNEKYADKSKEDL 93
 LEE MSNQHIEELNDQQIVRREKMA AL EQGIDPFGKRF-RTA S +L EKYADK+KE+L
 25 Sbjct: 1 LEENMSNQHIEELNDQQIVRREKMTALABQGIDPFGKRFDRTANSUELKEKYADKTKEEL 60

Query: 94 HDIEETATIAGRMTKRGKGKVGFQAHIQDREGQIQIYVRKDSVGEENEYEIFKKADLGDFL 153
 H++ ETA +AGRMTKRGKGKVGFQAH+QDREGQIQ+YVRKDSVGE+NYEIFKKADLGDF+
 Sbjct: 61 HELNETAIVAGRMTKRGKGKVGFQAHLDREGQIQLYVRKDSVGEDNYEIFKKADLGDFI 120

30 Query: 154 GVEGQVMRTDMGELSIAKATHITHLSKALRPLPEKFHGLTDIETIYRKRHLDLISNRDSFD 213
 GVEG+VMRTDMGELSIAKAT +THLSK+LRPLPEKFHGLTDIETIYRKRHLDLISNR+SFD
 Sbjct: 121 GVEGEVMRTDMGELSIAKATKLTHLSKSLRPLPEKFHGLTDIETIYRKRHLDLISNRRESFD 180

35 Query: 214 RFVTRSKIISEIRRFMDSNGFLEVETPVHLNEAGGASARPFI THHNAQDIDMVLRIATEL 273
 RFVTRSK+ISEIRR++D FLEVETPVHLNEAGGA+ARPF+THHNAQ+IDMVLRIATEL
 Sbjct: 181 RFVTRSKMISEIRRYLDGLDFLEVETPVHLNEAGGAAARPFTVHHNAQNIIDMVLRIATEL 240

40 Query: 274 HLKRLLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEAYQAYADYQDIMDLTEGIIQHVT 333
 HLKRLLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIE YQAYADY DIM+LTEGIIQH
 Sbjct: 241 HLKRLLIVGGMERVYEIGRIFRNEGMDATHNPEFTSIEVYQAYADYLDIMNLTEGIIQHAA 300

45 Query: 334 KTVKGDPINYQGTEIKINEPFKRVHMVDAVKEITGIDFWKEMTLEEAQALAQEKNVPLE 393
 K V+GDPPI+YQGTEI+INEPFKRVHMVA+KE+TG DFW EMT+EEA ALA+EK VPLE
 Sbjct: 301 KAVRGDGPIDYQGTEIRINEPFKRVHMVDAIKEVTGADFVPEMTVEEAIALAKEKQVPLE 360

50 Query: 394 KHFTTVGHIIINAFEEFVEDTLIQPTFVFGHPVEVSPLAKKNNDTDPRTDRFELFIMTKE 453
 KHF +VGHIINAFEEFVE+TL+QPTFVFGHPVEVSPLAKKN D RFTDRFELFIMTKE
 Sbjct: 361 KHFISVGHIINAFEEFVEETLVQPTFVFGHPVEVSPLAKKNPEDTRFTDRFELFIMTKE 420

Query: 454 YANAFTELNDPIDQLSRFEAQASAKELGDDATGVVDYDYEVALEYGMPPPTGGLGIGIDRL 513
 YANAFTELNDPIDQLSRFEAQ AKELGDDATG+DYD+VEALEYGMPPPTGGLGIGIDRL
 Sbjct: 421 YANAFTELNDPIDQLSRFEAQAKELGDDATGIDYDFVVALEYGMPPPTGGLGIGIDRL 480

55 Query: 514 CMLLTDTTTIIRDVLLFPTMK 533
 CMLLT+TTTIIRDVLLFPTMK
 Sbjct: 481 CMLLTNTTTIIRDVLLFPTMK 500

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 60 vaccines or diagnostics.

-421-

Example 331

A DNA sequence (GBSx0361) was identified in *S.agalactiae* <SEQ ID 1075> which encodes the amino acid sequence <SEQ ID 1076>. This protein is predicted to be 6,7-dimethyl-8-ribityllumazine synthase (ribH). Analysis of this protein sequence reveals the following:

5 Possible site: 34
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.1042 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14257 GB:Z99116 riboflavin synthase (beta subunit) [Bacillus subtilis]
 15 Identities = 103/151 (68%), Positives = 120/151 (79%)

Query: 1 MTIIEGQLVANEMKIGIVVSRFNELITSKLLSGAVDGLLRHGVSEEDIDIVWVPGAFEIP 60
 M II+G LV +KIGIVV RFN+ ITSKLLSGA D LLRHGV DID+ WVPGAFEIP
 Sbjct: 1 MNIIQGNLVGTGLKIGIVVGRFNDIFTSKLLSGAEDALLRHGVDTNDIDVAWVPGAFEIP 60
 20 Query: 61 YMARMKALYKDYDAIIICLGVVIKGSTDHYDVVCNEVTKGIGHLNSQSDIPHIFGVLTIDN 120
 + A+KMA K YDAII LG VI+G+T HYDYVCNE KGI + + +P IFG++TT+N
 Sbjct: 61 FAAKKMAETKKYDAIIITLGTIVRGATTHYDVVCNEAKGIAQAANTTGVPVIFGIVITEN 120
 25 Query: 121 IEQAIERAGTKAGNKGYDCALSAIEMVNLDK 151
 IEQAIERAGTKAGNKG DCA+SAIEM NL++
 Sbjct: 121 IEQAIERAGTKAGNKGVDCAVSAIEMANLNR 151

No corresponding DNA sequence was identified in *S.pyogenes*.

30 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 332

A DNA sequence (GBSx0362) was identified in *S.agalactiae* <SEQ ID 1077> which encodes the amino acid sequence <SEQ ID 1078>. This protein is predicted to be GTP cyclohydrolase ii (ribA/B). Analysis of this protein sequence reveals the following:

Possible site: 20
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 40 bacterial cytoplasm --- Certainty=0.1918 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9725> which encodes amino acid sequence <SEQ ID 9726> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA86524 GB:U27202 GTP cyclohydrolase II/
 3,4-dihydroxy-2-butanone-4-phosphate synthase
 [Actinobacillus pleuropneumoniae]
 50 Identities = 230/395 (58%), Positives = 307/395 (77%)

Query: 19 FSPIKKLLQDIKSGKMVVLMDDENRENEGDLICAAEMVTKESINFMAFKGKGLICLPLSN 78
 FS ++ ++ I+ GK++++ DDE+RENEGD ICAAEE T E+INFMA +GKGLIC P+S
 Sbjct: 6 FSKVEDAAIEAIRQGKIIILVTDEDRENEGDFICAAEFATPENINFMATYGKGLICTPIST 65

-422-

Query: 79 YYAEKLELAQMASHNTDNHETAFITISIDHLSTSTGISAEDRALTAKMVANDSSKAKDFRR 138
A+KL M + N DNHETAFIT+S+DH+ T TGISA +R++TA + +D++KA DFRR
Sbjct: 66 EIAKQLNFHPMVAVNQDNHETAFITVSDHIDTGTGISAVERSITAMKIVDDNAKATDFRR 125

5

Query: 139 PGHLFPLLAKEGGVLRNNGHTEATVDLCRLAGLKECGLCCEIMAEDGSMMRKDELLAFAQ 198
PGH+FPL+AEGGVVL RNNGHTEATVDL RLAGLK GLCCEIMA+DG+MM +L FA
Sbjct: 126 PGHMFPLIAMEGGVLRNNGHTEATVDLARLAGLKHAAGLCCBIMADDGTMMTMPDLQKFAV 185

10

Query: 199 KHDIAITATIKQLQDYRRQEEGGVVREIEIQLPTQFGHFTAYGYSEVVANKEHVALVKGDI 258
+H++ TI+QLQ+YR++ + V + +++PT++G F A+ + EV++ KEHVALVKGD+
Sbjct: 186 EHNMPFITIQQLQEQYRRKHDLSLVKQISVVKMPTKYGEFMAHSFVEVISGKEHVALVKGDL 245

15

Query: 259 SSGEDVLCRLHSECLTDVFHSRLCDCGEQLANALQQIEAEGRGVLLYMRQEGRGIGLIN 318
+ GE VL R+HSECLTGDF S RDCG+Q A A+ QIE EGRGV+LY+RQEGRGIGLIN
Sbjct: 246 TDGEQVLARIHSECLTDGFSGQRCDGQQFAAAAMTQIEQEGRGVILYLRQEGRGIGLIN 305

20

Query: 319 KLKAYHLQEEGLDTLEANLALGFEGDERDYGVSAQLLKDLGINSTINLLNNPDKIQQLEA 378
KL+AY LQ++G+DT+EAN+ALGF+ DER+Y + AQ+ + LG+ SI LLTNNP KI+ L+
Sbjct: 306 KLRAYELQDKGMDTVEANVALGFKEDEREYYIGAQMFQQILGVKSIRLLTNNPAKIEGLKE 365

25

Query: 379 EGICVKNRVPLQVAVTAYDLNYLKTKEKGHLLD 413
+G+ + R P+ V D++YLK K+ KMGH+ +
Sbjct: 366 QGLNIVAREPIIVEPNKNDIDYLKVQIKMGHMFN 400

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 333

30 A DNA sequence (GBSx0363) was identified in *S.agalactiae* <SEQ ID 1079> which encodes the amino acid sequence <SEQ ID 1080>. This protein is predicted to be riboflavin synthase alpha chain (ribE). Analysis of this protein sequence reveals the following:

Possible site: 59
>>> Seems to have no N-terminal signal sequence
35 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.3517(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>

40 A related GBS nucleic acid sequence <SEQ ID 9723> which encodes amino acid sequence <SEQ ID 9724> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB05274 GB:AP001512 riboflavin synthase alpha subunit [Bacillus halodurans]
45 Identities = 98/216 (45%), Positives = 147/216 (67%), Gaps = 2/216 (0%)

Query: 1 MFTGIIEMGQVSIRNGIKSQLSIDAPKLVPPLRKGD\$AVANGVCLTVLDKSETAFIA 60
MFTGIIEMGQVSIRNGIKSQLSIDAPKLVPPLRKGD\$AVANGVCLTVLDKSETAFIA 60
Sbjct: 1 MFTGIIEDVGTIDAIQQTGEIAVMTITSKKIVSDVQLGDSIAVANGVCLTVTSFTDTQFTV 60

50

Query: 61 DVMPESMMRSLAALRLHSKVNLALRSDSLRLGGHFVLGHVDGVGKIEKIQKDDIAVRF 120
D+MPE++ TSL L S+VNLE A+ ++ R GGH V GHVDG+G I K ++ D AV +
Sbjct: 61 DLMPETVRATSLRLLSKGSRVNLERAMVANGRFGGGHIVSGHVDGIGTIRKKERKDNAVYY 120

55

Query: 121 SIDAPPSSIMSYIIIEKGSVALDGISLTIVVSFTEH\$FEVSVIPHTMAQTNLSLKKVGDLNI 180
+I+ S+ Y+I KGSVA+DG SLT+ ++ +F +S+IPHTM +T + LKK GD++NI
Sbjct: 121 TIEVSSSSLRRYMIHKGSVAVDGTSLTIFDVSDDKTFITISIIPHTMEETIIGLKKAGDIVNI 180

-423-

Query: 181 EVDVLGKYAEKFLAPTNRTNHTSSVMDWSFILSENGY 216
 E D++GKY E+F+ N + +FL+E+GY
 Sbjct: 181 ECDLIGKYIEQFVQQGKPVNNEGG--LTKAFLTEHGY 214

- 5 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 334

- A DNA sequence (GBSx0364) was identified in *S.agalactiae* <SEQ ID 1081> which encodes the amino acid sequence <SEQ ID 1082>. This protein is predicted to be riboflavin-specific deaminase (ribD). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.01    Transmembrane 307 - 323 ( 307 - 323)
15
----- Final Results -----
      bacterial membrane --- Certainty=0.1404 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
20
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA86522 GB:U27202 riboflavin-specific deaminase [Actinobacillus
      pleuropneumoniae]
      Identities = 182/353 (51%), Positives = 259/353 (72%)
25
Query: 6   DYMALALKEAEKGGMGFVAPNPLVGAVIVKDRIISKGYHKRFGDLHAERQAIAKNADEDIS 65
      DYM A+ A++G+G+ PNPLVG VIVK+ I+++GYH++ G HAER A+ + ED+S
      Sbjct: 51  DYMRAIALAKQGLGWTPNPVLVGCVIVKNGEIVAEGYHEKIGGWAERNAVLHKCEDLS 110
30
Query: 66  GSTLYVTLEPCCHVGKQPPCTEALIKSGIKVVVGSLDPNPLVSGKGIALLRKEGLNVEV 125
      G+T YVTLEPCCH G+ PPC++ LI+ GIKKV +GS DPNPLV+G+G LR+ G+ V
      Sbjct: 111 GATAYVTLEPCCHHGRTPPCSDLLIERGKKVFIGSSDPNPLVAGRGANQLRQAGVEVVE 170
35
Query: 126 GILREECDALNERFIFHMTYKOPFVYLKYAMTLDGKIATKTDGSKWISNEHSRQSVKLR 185
      G+L+EECDALN F ++ K+P+V +KYAMT DGKIAT +G+SKWI+ E +R VQ+ R
      Sbjct: 171 GLLKEECDALNPPIFFHYIQTKRPVLMKYAMTADGKIATGSGESKWITGESARARVQQTR 230
40
Query: 186 QKCSAIMVGINTVLADNPRLTCRIPKGEALVRIVCDSQLKIPLDSYLVSAKTIPTWIAT 245
      + SAIMVG++TVLADNP L R+P + VRIVCDSQL+ PLD LV++AK T IAT
      Sbjct: 231 HQYSAIMVGVDTVLADNPMLNSRMPNAKQPVRIVCDSQLRPLDCQLVQTKEYRTVIAT 290
45
Query: 246 CSDNLAQQQTLKEMGCRLIKVPRKDGLKDLKVLMTILGQEGIDSLLIEGGSSLHFSAK 305
      SD+L + + +G ++ ++ ++DL+ L+ LG+ IDSLL+EGGSSL+FSAL++
      Sbjct: 291 VSDDLQKIEQFRPLGVVDVLVCKARNKRVDLQDLLQKLGEAQIDSLLEGGSSLNFSALES 350
Query: 306 GIVNRRLIVFIAPKIIIGGLKAKTAISGEGLDWLNAQAFRVKDIELSRSMDSDVVIE 358
      GIVNR+ +IAPK++GG +AKT I GEG+ ++QA ++K + D++++
      Sbjct: 351 GIVNRVHCYIAPKLVGGKQAKTPIGGEGIQQIDQAVKLKLKSTELIGEDILLD 403
```

- 50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1083> which encodes the amino acid sequence <SEQ ID 1084>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.17    Transmembrane 88 - 104 ( 88 - 105)
55
----- Final Results -----
      bacterial membrane --- Certainty=0.1468 (Affirmative) < succ>
```

-424-

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:CAB11794 GB:Z99104 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 71/161 (44%), Positives = 109/161 (67%)

Query: 13 LEEQTYFMQEALKAEKSLQKAEIPIGCVIVKDGEIIGRGNAREESNQAIMHAEMMAIN 72
 + + +M+EA+KEA+K+ +K E+PIG V+V +GEII R HN RE ++I HAEM+ I+

10 Sbjct: 1 MTQDELYMKAEAIKEAKKAEEKGEVPIGAVLIVINGEIIARAHNLRETEQRSIAHAEMLVID 60

Query: 73 EANAHEGNWRLLDTTLFVTIEPCVMCSGAIGLARIPHVIYGASNQKFGGVDSLQILTDE 132
 EA G WRL TL+VT+EPC MC+GA+ L+R+ V++GA + K G +L +L +E

15 Sbjct: 61 EACKALGTRLEGATLYVTLEPCPMCAGAVVLSRVEKVVFAGDPKGCGSGTLMNLLQEE 120

Query: 133 RLNRVQVERGLLAADCANIMQTFFRQGRERKKIAKHLIKE 173
 R NH+ +V G+L +C ++ FFR+ R++KK A+ + E

Sbjct: 121 RFNHQAEVVSGVLEEECGGMLS AFFRELRKKKAAARKNLSE 161

20 An alignment of the GAS and GBS proteins is shown below:

Identities = 48/146 (32%), Positives = 71/146 (47%), Gaps = 21/146 (14%)

25 Query: 7 YMALALKEAEKGGMGFVAPNPLVGAVIVKDDRIISKGYHKRGD---LHAERQAIIKNADE 62
 +M ALKEAEK + A P +G VIVKD II +G++ R +HAE AI A+
 Sbjct: 19 FMQEALKAEKSLQ-KAEIPI-IGCVIVKDGEIIGRGNAREESNQAIMHAEMMAINEANA 76

Query: 63 D----ISGSTLVTLEPCCHVGKQPPCTEALIKSGIKKVVVGSLDPNPLVSGKGIAALLR 117
 + +TL+VT+EPC C+ A+ + I V+ G+ + +L

30 Sbjct: 77 HEGNWRLLDTTLFVTIEPCV-----MCSGAIGLARIPHVIYGASNQKFGGVDSLQILT 130

Query: 118 KEGLN----VEVGILREECDAILNERF 139
 E LN VE G+L +C + + F

Sbjct: 131 DERLNHRVQVERGLLAADCANIMQT 156

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 335

A DNA sequence (GBSx0365) was identified in *S.agalactiae* <SEQ ID 1085> which encodes the amino acid sequence <SEQ ID 1086>. This protein is predicted to be Nramp metal ion transporter. Analysis of this 40 protein sequence reveals the following:

Possible site: 43

>>> Seems to have no N-terminal signal sequence

45	INTEGRAL	Likelihood = -11.89	Transmembrane	169 - 185 (160 - 191)
	INTEGRAL	Likelihood = -11.09	Transmembrane	140 - 156 (128 - 165)
	INTEGRAL	Likelihood = -6.85	Transmembrane	359 - 375 (354 - 379)
	INTEGRAL	Likelihood = -6.48	Transmembrane	269 - 285 (263 - 287)
	INTEGRAL	Likelihood = -6.16	Transmembrane	426 - 442 (423 - 445)
	INTEGRAL	Likelihood = -5.57	Transmembrane	62 - 78 (58 - 80)
50	INTEGRAL	Likelihood = -4.94	Transmembrane	107 - 123 (103 - 127)
	INTEGRAL	Likelihood = -4.46	Transmembrane	391 - 407 (389 - 408)
	INTEGRAL	Likelihood = -4.35	Transmembrane	310 - 326 (307 - 328)

----- Final Results -----

55 bacterial membrane --- Certainty=0.5755 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF83825 GB:AE003939 manganese transport protein [Xylella

-425-

fastidiosa]

Identities = 192/436 (44%), Positives = 274/436 (62%), Gaps = 14/436 (3%)

Query: 10 SLSEVNQSVEVPHNSSFNWNTLRAFLGPGALAVGYMDPGNWITSVIGGATYRYLLLFFVVL 69

SL E++ SV V + L AFLGPG +V+VGYMDPGNW T + GG+ + Y+LL V+L

Sbjct: 39 SLGEMHASVAVSRGRGHNGFRLLAFLGPGYMSVGYMDPGNWATGLAGGSRGYMLLSVIL 98

Query: 70 VSSLMAMQLQQMAGKLGIVTRQDLAQATASRLPKPLRYYLLFIIIELALIATDLAEVIGSA 129

+S++MA+ LQ +A +LGI + DLAQA +R + L+++ ELA+IA DLAEVIG+A

Sbjct: 99 LSNVMAIVLQALAARLGIASDMDLAQACRARYSRGTTIALWVVCCELIAACDLAEVIGTA 158

Query: 130 IALHLLFGWPULLSIMITILDVFLLLLLMKLGQKIEAFVSVLILTILIIIFTYLVVLSQP 189

IAL+LL G P++ ++IT +DV L+LLM G + +EAFT L+L I F +VL+ P

Sbjct: 159 IALNLLGVPIIWGVVITAVDVLVLLLMHRGFRALEAFVIALLLVIFGCFVVQIVLAAP 218

Query: 190 DLDAMFKGFLPHHELFNISHEGKNSPPLTLALGIIGATVMPHNLYLHSSLSQTRRVDYHNK 249

L + GF+P ++ L LA+GI+GATVMPHNLYLHSS+ QTR

Sbjct: 219 PLQEVLGGFVPRWQVV----ADPQALYLAIGIVGATVMPHNLYLHSSIVQTRAYP-RTP 272

Query: 250 SSIKKAVRFMTLDSNIQLSLAFVVNSLLLVLGASLFY-G-HANDISAFSQMYLALSDDKTIT 308

+ A+R+ DS + L LA +N+ +L+L A++F+ H D+ Q Y L+

Sbjct: 273 VGRRSALRWAVADSTLALMLALFINASILILAAAVFHQAHHFDVVEEQAYQLLAPVLGV 332

Query: 309 GAVASSFLSTLFAVALLASGQNSTITGTLTGQIVMEGFLHFKLPQWLIRLCTRLLTLPI 368

G A TLFA ALLASG NST+T TL GQIVMEGFL +L WL R+ TR L ++P+

Sbjct: 333 GVAA----TLFATALLASGINSTVTATLAGQIVMEGFLRLRLRPWLRRVLTRGLAIPV 387

Query: 369 FVIALLVGGEENTLDQLIVYSQVFLSLALPFSIFPLIYFTSQKSIMGEHANAKWNTYLAY 428

V+ L G E +L++ SQV LS+ LPF++ PL+ + + +MG +W +A+

Sbjct: 388 IVVVALYG--EQGTGRLLLLSQVILSMQLPFAVIPLLRCVADRKGALVAPRWLMMVVAW 445

Query: 429 LVAITLTLNLKIMD 444

L+A ++ +LN+KL+ D

Sbjct: 446 LIAGVIVVNLNVKLLGD 461

35

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 336

40 A DNA sequence (GBSx0366) was identified in *S.agalactiae* <SEQ ID 1087> which encodes the amino acid sequence <SEQ ID 1088>. Analysis of this protein sequence reveals the following:

Possible site: 54

>>> Seems to have a cleavable N-term signal seq.

INTEGRAL Likelihood = -14.12 Transmembrane 113 - 129 (98 - 132)

INTEGRAL Likelihood = -12.15 Transmembrane 228 - 244 (220 - 249)

INTEGRAL Likelihood = -10.83 Transmembrane 175 - 191 (167 - 195)

INTEGRAL Likelihood = -5.04 Transmembrane 57 - 73 (55 - 75)

INTEGRAL Likelihood = -3.93 Transmembrane 146 - 162 (142 - 166)

INTEGRAL Likelihood = -1.38 Transmembrane 199 - 215 (199 - 215)

INTEGRAL Likelihood = -0.32 Transmembrane 82 - 98 (82 - 98)

----- Final Results -----

bacterial membrane --- Certainty=0.6647(Affirmative) < succ>

bacterial outside --- Certainty=0.0000(Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

55

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAF11325 GB:AE002018 hypothetical protein [Deinococcus radiodurans]

Identities = 63/215 (29%), Positives = 108/215 (49%), Gaps = 13/215 (6%)

60

Query: 11 LLLVFILTIIVNYLSATGFLTGNSQKSLSDRYQTLLTPAPLAFSIWSVIYL-LTFLVILR 69

-426-

```

      LL +LT++VNYLS   L GNS +SDR      TPA L F++W I+L L    +
Sbjct: 10 LLAATVLTLVVNYLSNALPLFGNSNAEVSDRLPNAFTPAGLTFTVWGP1FLGLLVFAVYQ 69

      Query: 70 AIFSKSQSYQDNFASIFPYFLGLLLVNNIWTVFFTSNLIGLSTIIIFAYCILLV-IIIKI 128
      A+ ++ + D +P+ LG LL N W + F S IGLS +I+ A +LV + +
Sbjct: 70 ALPAQRGARLDRL--FWPFLLGNLL-NVAWLLAFQSLNIGLSVVIMALLAVLVRLYLSV 126

      Query: 129 LS---KNKSKLLLRTFGIHAGWLLVASLVLNVNLAVYLVKI---DFNYPLPKVYIAIIALI 181
      S + + L++ ++ W+ VA++ N+ +LV F V+ A++ ++
Sbjct: 127 RSLPPQGAERWTLQLPVSLYLAWSVATIANITAFLVSAGVTQSFLGIAGPVWSALLVV 186

      Query: 182 FITVLSLYLARVLQNAYLILSVFWAFLMVFKAHLE 216
      + +L R A+ + + WA+ V+ A E
Sbjct: 187 AAAIGVFFLWRFRDYAFAAV-LLWAFYGVYVARPE 220

```

15 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 337

20 A DNA sequence (GBSx0367) was identified in *S.agalactiae* <SEQ ID 1089> which encodes the amino acid sequence <SEQ ID 1090>. Analysis of this protein sequence reveals the following:

```

Possible site: 36
>>> Seems to have no N-terminal signal sequence

25 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3401(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

30 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAC65352 GB:AE001215 T. pallidum predicted coding region
      TP0352 [Treponema pallidum]
      Identities = 28/64 (43%), Positives = 41/64 (63%)

```

35 Query: 3 EFTFEIVEKLLVLSENEKGWTKELENRVSFNGAPAKFDLRTWSPDHTKMGKGITLSNEEFK 62
 +F+E+ LS + GW+ EL +S+NG P K+D+R WSPD +KMGKG+TL+ E
Sbjct: 12 DFHYEVTRNWGTLSSTSGNGWSLELKSISWNGRPEKYDIRAWSPDKSKGKVTLTRAEIF 71

40 Query: 63 VILD 66
 + D
Sbjct: 72 ALRD 75

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1091> which encodes the amino acid sequence <SEQ ID 1092>. Analysis of this protein sequence reveals the following:

```

45 Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
      bacterial cytoplasm --- Certainty=0.4021(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

An alignment of the GAS and GBS proteins is shown below:

```

55 Identities = 59/70 (84%), Positives = 64/70 (91%)

Query: 1 MSEFTFEIVEKLLVLSENEKGWTKELENRVSFNGAPAKFDLRTWSPDHTKMGKGITLSNEE 60
      M+EFTF I E LL LSEN+KGWTKELENRVSFNGA AK+D+RTWSPDHTKMGKGITL+NEE
Sbjct: 1 MAEFTFNIEEHLTLSENDKGWTKELENRVSFNGAEAKWDIRTWSPDHTKMGKGITLTNEE 60

```

Query: 61 FKVILDRAFRK 70
 FK ILDAFRK
 Sbjct: 61 FKTILDAFRK 70

5

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 338

A DNA sequence (GBSx0368) was identified in *S.agalactiae* <SEQ ID 1093> which encodes the amino acid sequence <SEQ ID 1094>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -2.66   Transmembrane  92 - 108 ( 92 - 110)

15 ----- Final Results -----
    bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

20 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14676 GB:Z99117 similar to protease [Bacillus subtilis]
  Identities = 201/407 (49%), Positives = 277/407 (67%), Gaps = 2/407 (0%)
```

25 Query: 4 VKKRPEVLSPAGTLEKLKVAIDYGADAVFVGQQAYGLRSRAGNFSMEELQEGINYAHARD 63
 + K+PE+L+PAG LEKLK+A+ YGADAVF+GGQ YGLRS A NF++EE+ EG+ +A
 Sbjct: 18 ITKKPELLAPAGNLEKLKIAVHYGADAVFIGGQEYGLRSNADNFTIEEIAEGVEFAKKYG 77

30 Query: 64 AKVYVAANMVTHEGNELGAGPWRELDMGLDAVIVSDPALIVICATEAPGLEIHLSTQA 123
 AK+YV N+ H N G + + L D + +IV+DP +I C AP +E+HLSTQ
 Sbjct: 78 AKIYVTTNIFAHNENMDGLEDYLKALGDANVAGIIVADPLIETCRRVAPNVEVHLSTQQ 137

35 Query: 124 SSTNYETFEFWKEMGLTRVVLAREVTMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLS 183
 S +N++ +FWKE GL RVVLARE + E+ E+++ D+EIE+F+HGAMCI+YSGRCVLS
 Sbjct: 138 SLSNWKAQFWKEEGLDRVVLARETSALLEIREMKEVDIEIESFIHGAMCIAYSGRCVLS 197

40 Query: 184 NHMSHRDANRGCGSQSCRWKYDLYDMPFGQERQSLKGIEIPEPFSMSAVDMCMIEHIPDMI 243
 NHM+ RD+NRGGC QSCRW YDLY G +L GE PF+MS D+ +IE IP MI
 Sbjct: 198 NHMTARDSNRGGCCQSCRWDYDLYQTD-GANAVALYGEEDAPFAMSPKDLKLIESIPKMI 256

45 Query: 244 ENGVDSLKIEGRMKSIHYVSTVNCYKAADVDAYMESPEAFEAIKEDELWVKVAQRELA 303
 E G+DSLKIEGRMKSIHYV+TV + Y+ +DAY PE F I+++ ++EL K A R+ A
 Sbjct: 257 EMGIDSLSKIEGRMKSIHYVATVSVYRKVIDAYCADPENF-VIQKEWLEELDKCANRDTA 315

50 Query: 304 TGFYYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQRNVIMEGDRVEFYGPG 363
 T F+ TP EQ+FG K Y FVG V+++D T++QRN +GD VEF+GP
 Sbjct: 316 TAFFEGTPGYEEQMFGEHAKKTTYDFVGLVINYDEDTQMVTLQQRNFFKGDEVEFFGPE 375

Query: 364 FRHFECFIDGLRDAEGNPKIDRAPNPMLLTITLPNPVKKGDMIRACK 410
 +F I+ + D +GN++D A +P+++ L + +M+R K
 Sbjct: 376 IENFTHTIETIWIWEDGNELDAARHPLQIVFKLDKKIYPSNNMMRKKGK 422

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1095> which encodes the amino acid sequence <SEQ ID 1096>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
    INTEGRAL      Likelihood = -2.66   Transmembrane  92 - 108 ( 92 - 110)

55 ----- Final Results -----
    bacterial membrane --- Certainty=0.2062(Affirmative) < succ>
    bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

60

-428-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAB04993 GB:AP001511 protease [Bacillus halodurans]
 5 Identities = 201/403 (49%), Positives = 280/403 (68%), Gaps = 4/403 (0%)

Query: 6 KRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAHARGAK 65
 K+PE+L+PAG+LEKLKVAI YGADAV++GGQ +GLRS A NFS+EE++EG+++A+ GAK
 Sbjct: 17 KKPELLAPAGSLEKLKVAIHYGADAVYIGGQEFGLRSNADNFSIEEMREGVEFANKYGAK 76

10 Query: 66 VYVAANMVTHEGNEIGAGEWFRLRDMGLDAVIVSDPALIVICSTEAPGLEIHLSTQASS 125
 VYV N+ H N G E+ L+++G+ +IV+DP +I C AP +E+HLSTQ S
 Sbjct: 77 VYVTNTNIYAHNENMDGLEEYLSALQEVGVTGIIADPLIETCKRVAPKVEVHLSTQQL 136

15 Query: 126 TNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGRCVLSNH 185
 +N+ +FWK GL RVVLAREV + E+ E++K D+EIE FVHGAMCISYSGRCVLSNH
 Sbjct: 137 SNWLAVKFWKEEGLHRVVLAREVGLEEMLEMKKHVVDIEIETFVHGAMCISYSGRCVLSNH 196

20 Query: 186 MSHRDANRGGSQSQRWKYDLYDMPFGGE-RRSLKGEIPEPDYSMSSVDMCMIDHIPDLIE 244
 M+ RD+NRGGC QSCRW YDLY+ E +G++P Y+MS D+ +I IP LIE
 Sbjct: 197 MTARDSNRGGCCQSQRWDYDLYEQQDSAEIPLFAEGDVPA--YTMSPKDNLNIQAIPQIE 254

25 Query: 245 NGVDSLKIEGRMKSIHYVSTVTNCYKAAVGAYMESPEAFYAIKEELIDEWLKVQAQRELAT 304
 G+DSLK+EGRMKSIHYV+TVT+ Y+ + AY P+ F IK E ++EL K A R+ A
 Sbjct: 255 AGIDSLKVEGRMKSIHYVATVTSVYRKVIDAYCSDPDNF-KIKREWLEELEKCANRDFAP 313

30 Query: 305 GFYYGIPTENEQLFGARRKIPQYKFVGEVVAFDSSMTATIRQRNVIMEGDRIECYGGPF 364
 F+ G PT EQ++G K +Y FVG V+ ++ + T++QRN +GD +E +GP
 Sbjct: 314 QFFEGTPTYKEQMYGIHPKRTKYDFVGLVLDYNEKTGIVTLQQRNHFQQGDEVEFFGPEI 373

35 Query: 365 RHFETVVVKDLHDADGQKIDRAPNPMEELLTISLPREVKGDMIR 407
 F V+ + D DG ++D A +P++++ + ++V P +M+R
 Sbjct: 374 NRFTQTVEKIWDEDGNELDAARHPLQIVKFKVDQKVYPQNMMR 416

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 386/427 (90%), Positives = 404/427 (94%)

Query: 1 MSNVKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGINYAH 60
 MS++KKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGI+YAH
 40 Sbjct: 1 MSHMKKRPEVLSPAGTLEKLKVAIDYGADAVFVGGQAYGLRSRAGNFSMEELQEGIDYAH 60

Query: 61 ARDAKVVVAANMVTHEGNEIAGPWFRRLRDMGLDAVIVSDPALIVICATEAPGLEIHL 120
 AR AKVVVAANMVTHEGNE+GAG WFR+LRDMGLDAVIVSDPALIVIC+TEAPGLEIHL
 Sbjct: 61 ARGAKVVVAANMVTHEGNEIGAGEWFRLRDMGLDAVIVSDPALIVICSTEAPGLEIHL 120

45 Query: 121 TQASSTNYETFEFWKEMGLTRVVLAREVMAELAEIRKRTDVEIEAFVHGAMCISYSGR 180
 TQASSTNYETFEFWK MGLTRVVLAREV MAELAEIRKRTDVEIEAFVHGAMCISYSGR
 Sbjct: 121 TQASSTNYETFEFWKAMGLTRVVLAREVNMAELAEIRKRTDVEIEAFVHGAMCISYSGR 180

50 Query: 181 VLSNHNMSHRDANRGGSQSQRWKYDLYDMPFGQERQSLKGEIPEPFMSA VDMCMIEHIP 240
 VLSNHNMSHRDANRGGSQSQRWKYDLYDMPFG ER+SLKGEIPE +SMS+VDMCM+HIP
 Sbjct: 181 VLSNHNMSHRDANRGGSQSQRWKYDLYDMPFGERRSLKGEIPEPDYSMSSVDMCMIDHIP 240

55 Query: 241 DMIENGVDLSKIEGRMKSIHYVSTVTNCYKAAVDAYMESPEAFEAIIKEDLIDELWKVAQR 300
 D+IENGVDLSKIEGRMKSIHYVSTVTNCYKA AV AYMESPEAF AIKE+LIDELWKVAQR
 Sbjct: 241 DLIENGVDLSKIEGRMKSIHYVSTVTNCYKAAVGAYMESPEAFYAIKEELIDELWKVAQR 300

60 Query: 301 ELATGFYYHTPTENEQLFGARRKIPQYKFVGEVVSFDNAKMEATIRQRNVIMEGDRVEFY 360
 ELATGFYY PTENEQLFGARRKIPQYKFVGEVV+FD+A M ATIRQRNVIMEGDR+E Y
 Sbjct: 301 ELATGFYYGIPTENEQLFGARRKIPQYKFVGEVVAFDSSMTATIRQRNVIMEGDRIECY 360

65 Query: 361 GPGFRHFECFIDGLRDAEGNKIDRAPNPMELLTITLPNPVKKGDMIRACKEGLVNL 420
 GPGFRHFEC + L DA+G KIDRAPNPMELLTI+LP VK GDMIRACKEGLVNLQD
 Sbjct: 361 GPGFRHFETVVVKDLHDADGQKIDRAPNPMELLTISLPREVKGDMIRACKEGLVNLQKD 420

Query: 421 GTSKTVR 427

GTSKTVR
 Sbjct: 421 GTSKTVR 427

SEQ ID 1094 (GBS385) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 69 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 72 (lane 7; MW 75.7kDa).

The GBS385-GST fusion product was purified (Figure 213, lane 7) and used to immunise mice. The resulting antiserum was used for FACS (Figure 312), which confirmed that the protein is immunoaccessible on GBS bacteria.

10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 339

A DNA sequence (GBSx0369) was identified in *S.agalactiae* <SEQ ID 1097> which encodes the amino acid sequence <SEQ ID 1098>. This protein is predicted to be collagenase. Analysis of this protein 15 sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
20      bacterial cytoplasm --- Certainty=0.2208 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
25      >GP:CAB14677 GB:Z99117 similar to protease [Bacillus subtilis]
          Identities = 92/304 (30%), Positives = 161/304 (52%), Gaps = 5/304 (1%)

          Query: 1 MEKIILTATAESIEQVKQLLAIGIDRITYVGEENYGLRLPHSFSDDELREIAKLVHDAGKE 60
                  M+K L T S + L+ G VGE+ YGLRL FS +++ + ++ H G +
          30      Sbjct: 1 MKKPELLVTPTSTADILPLIQAGATAFLVGEQRYGLRLAGEFSREDVTKAVERIAHKEGAK 60

          Query: 61 LTVACNALMHQEMMDNIKPFLLEMKEINVVDYLVVGDAGVFYINKRDGYNFKLTYDTSVFV 120
                  + VA NA+ H + + +L + E VD V GD V + + KL + T
          35      Sbjct: 61 VYVAVNAIFHNDKVGELEYLAFLAEAGVDAAVFGDPAVLMAARESAPDLKLHWSTETTG 120

          Query: 121 TSSRQVNFWCGOHGAVETVLAREIPSEELFKMSENLEFPAAEILVY GASVIHHSKRPLLQNY 180
                  T+ N+WG+ GA +VLARE+ + + ++ EN E EI V+G + + SKR L+ NY
          Sbjct: 121 TNYYTCNYWGRKGAAARSVLARELNMDSIDEIKEVANEVEIEIQVHGMTCMFQSRSILIGNY 180

          Query: 181 YNF---THITDEKTRERGLFLAEPGDPESHYSIYEDKHGTHIFINNDINMMTKVTELVEH 237
                  + + + K +E G+FL + + ++ Y I+ED++GTHI ND+ ++ ++ EL++
          40      Sbjct: 181 FEYQGKVMMDIERKKKESGMFLHDK-ERDNKYPIFEDENGTHIMSPNDVCIIIDELEELIDA 239

          Query: 238 HFTHWKLDGIYCPGDNFVAIAEIFVETARL-IENGTFQDQAFLFDERIRKLHPKGRGLD 296
                  +K+DG+ + + + ++ E L +EN + + + ERI + P R +D
          45      Sbjct: 240 GIDSKFKIDGVLMPEYLIEVTMYREAIIDLCAVENRDEYEAKKEDWIERIESIQPVNRKID 299

          Query: 297 TGFY 300
                  TGF+
          50      Sbjct: 300 TGFF 303
```

A related GBS nucleic acid sequence <SEQ ID 10949> which encodes amino acid sequence <SEQ ID 10950> was also identified.

-430-

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1099> which encodes the amino acid sequence <SEQ ID 1100>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

5 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1716 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

10 An alignment of the GAS and GBS proteins is shown below:

Identities = 245/308 (79%), Positives = 273/308 (88%)

15 Query: 1 MEKIIILTATAESIEQVKOLLAILGIDRIYVGEENYGLRLPHFSDDDELREIAKLVHDAGKE 60
 MEKII+TATAESIEQVK LLA G+DRIYVGE NYGLRLPH+FS DELR+IAKLVHDAGKE

Sbjct: 1 MEKIIITATAESIEQVKALLAAGVDRIYVGEANYGLRLPHNFSYDELRQIAKLVHDAGKE 60

20 Query: 61 LTVACNALMHQEMMDNIKPFLLEMKEINV DYL VVGDAGVFYINKRDGYNFKLIYDT SVFV 120
 LTVACNALMHQ+MMD IKPFL+LM EI VDYL VVGDAGVFY+N KRDGYNFKLIYDT SVFV

Sbjct: 61 LTVACNALMHQDMMDQIKPFLDLMIEIAV DYL VVGDAGVFYVN KRDGYNFKLIYDT SVFV 120

25 Query: 121 TSSRQVNFWGQHGAVETVLAREI PSEELFKMSENLEFP A E I L V Y G A S V I H H S K R P L L Q N Y 180
 TSSRQVNFWGQHGAVE+V L A R E I P S E L F M I E I A V D Y L V V G D A G V F Y V N K R D G Y N F K L I Y D T S V F V

Sbjct: 121 TSSRQVNFWGQHGAVESV L A R E I P S A E L F T L A E N L E F P A E V L V Y G A S V I H H S K R P L L E N Y 180

30 Query: 181 YNFTHITDEKTRERGLFLAEPGDPE SHYSIYEDKHGTHIFINNDI NMMTKVTELVEHHFT 240
 Y+FT I DE +RERGLFLAEPGD SHYSIYED HGTHI FINNDI+MM+K+ EL H T

Sbjct: 181 YHFTKIDDEVS RERGLFLAEPGDASSHYSIYEDNHGTHIFINNDIDMMSKLGE LYAHGLT 240

35 Query: 241 HWKLDGIYCPGDNFVAIAE IFVETARLIENGFTQDQ AFL FDERIRKLHPKG RGL DTGFY 300
 HWKLDGIYCPGD+FVAI ++F++ L+E G FTQ++A D+ + HP GRGL DTGFY

Sbjct: 241 HWKLDGIYCPGDDFVAITKLFQAKTLLEAGQFTQEEAEKL DQAVHAHHPAGRGL DTGFY 300

Query: 301 DFDPSTVK 308
 +FDP TVK

Sbjct: 301 EFDPKTIVK 308

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 340

A DNA sequence (GBSx0371) was identified in *S.agalactiae* <SEQ ID 1101> which encodes the amino acid sequence <SEQ ID 1102>. This protein is predicted to be cDNA EST yk542c12.5 comes from this gene. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have a cleavable N-term signal seq.

55 ----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAD15622 GB:U75480 unknown [Streptococcus mutans]

Identities = 69/152 (45%), Positives = 101/152 (66%), Gaps = 12/152 (7%)

55 Query: 1 MSKLFKTLVISAASGAAAAYFLTTKKGKELRKNAEKFYGEYKENPEEYHQIAKDKASEYS 60
 MSK KT +I A +GAAAAYFL+T KGK+ +K + + +YKENP+EYHQ A DK +EY

Sbjct: 1 MSKFLKTAIIAGTGAAAAYFLSTDKGKQFKKKIHQTFTDYKENPKEYHQYAADKVNEYK 60

-431-

Query: 61 NLAVDTFKDYKGKFESGELETTEDIVSAVKEKSGEVVDFANDFVNQAKSKFSDEDTAKKED 120
 ++AV +FKDYK KFE+GELT ++I+S+VKEK+ + FAN ++Q K + T +K +
 Sbjct: 61 DVAHSFKDKFETGELETKDNISSVKEASQAGKFANSKLSQVKDHIA--QTVEKAE 118

5 Query: 121 KAP-----ETKVVEDIVIDYKENTEDKE 142
 + +V+DIVIDY+ + K+
 Sbjct: 119 ASTNDAGIPLGEMKAQVDDIVIDYQAEEEKTKK 150

10 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1103> which encodes the amino acid sequence <SEQ ID 1104>. Analysis of this protein sequence reveals the following:

Possible site: 26
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -1.81 Transmembrane 15 - 31 (14 - 31)

15 ----- Final Results -----

bacterial membrane --- Certainty=0.1723 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

20 A related sequence was also identified in GAS <SEQ ID 9117> which encodes the amino acid sequence <SEQ ID 9118>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 19
 >>> Seems to have a cleavable N-term signal seq.

25 ----- Final Results -----

bacterial outside --- Certainty= 0.300 (Affirmative) < succ>
 bacterial membrane --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

30 An alignment of the GAS and GBS proteins is shown below:

Identities = 69/140 (49%), Positives = 91/140 (64%), Gaps = 8/140 (5%)

35 Query: 1 MSKLFKTLVISAASGAAAAYFLTTKKKGKELRKNAEKFYGEYKENPEEYHQIAKDKASEYS 60
 M+K FK LVI A SG AAAYFL+T+KGK L+ AEK Y YKE+P++YHQ AK+K SEYS
 Sbjct: 8 MNKSFKNLVIGAVSGVAAAYFLSTEKGKALKNRAEKAYQAYKESPDDYHQFAKEKGSEYS 67

Query: 61 NLAVDTFKDYKGKFESGELETTEDIVSAVKEKSGEVVDFANDFVNQAKSKFS-EDTAKKE 119

40 +LA DTF D K K SG+LT ED++ +K+K+ FV + K ++ E K++
 Sbjct: 68 HLARDTFYDVKDKLASGDLTKEDMLDLLKDRT-----TAFVQKTKETLAEVEAKEKQD 120

Query: 120 DKAPETKVVEDIVIDYKENTE 139

D + EDI+IDY E E
 Sbjct: 121 DVIIDLNEEDIIDYTEQDE 140

45 SEQ ID 1102 (GBS164) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 30 (lane 4; MW 17.4kDa).

The GBS164-His fusion product was purified (Figure 115A; see also Figure 200, lane 4) and used to immunise mice (lane 1+2+3 product; 20µg/mouse). The resulting antiserum was used for Western blot, FACS (Figure 115B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

-432-

Example 341

A DNA sequence (GBSx0372) was identified in *S.agalactiae* <SEQ ID 1105> which encodes the amino acid sequence <SEQ ID 1106>. Analysis of this protein sequence reveals the following:

```

Possible site: 19
5    >>> Seems to have an uncleavable N-term signal seq
        INTEGRAL      Likelihood =-16.93   Transmembrane   6 - 22 ( 1 - 31)

----- Final Results -----
10      bacterial membrane --- Certainty=0.7771(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
15      Identities = 88/129 (68%), Positives = 112/129 (86%)

Query: 1 MIEIAVLIIIAIAFVVVLVLGILFVLKKVSETIEETKQTIKVLTSDVNVTLYQTNEILAKAN 60
        M EIA+LI+AIAF VLV+ ++ +L+K+S+T++E++QT+K+LTSDVNVTLYQTNE+LAKAN
Sbjct: 1 MWEIALLIVIAIAFAVLVIYLILLLRKISDTVDESRQTLKILTSVDVNVTLYQTNELLAKAN 60

20      Query: 61 VLVDDVNGKVSTIDPLFVAIALSESVDLNLQARHIGOKASSATSSVTKAGSALAIKGKA 120
        VLV+DVNGKV TIDPLF AIADLS SVSDLN QAR+ G+K +T++V KAG+A GK
Sbjct: 61 VLVEDVNGKVEITDPLFTAIAADLSVSVDLNRQARYFGKKTRKSTANVGKAGAAAYTFGKV 120

25      Query: 121 ASKIFRKKG 129
          ASK+FRKKG
Sbjct: 121 ASKLFRKKG 129
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1107> which encodes the amino acid sequence <SEQ ID 1108>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have a cleavable N-term signal seq.
        INTEGRAL      Likelihood = -0.85   Transmembrane   18 - 34 ( 17 - 34)

----- Final Results -----
35      bacterial membrane --- Certainty=0.1341(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

40 The protein has homology with the following sequences in the databases:

```

>GP:AAD15621 GB:U75480 unknown [Streptococcus mutans]
45      Identities = 83/128 (64%), Positives = 110/128 (85%)

Query: 6 ISLMIIIALAFVALVFLIIVLKKVSETIDEAKKTISVLTSVDVNVTLHQTNIDILAKANILV 65
        I+L+I+A+AF LVI+LI++L+K+S+T+DE++T+ +LTSDVNVTL+QTN++LAKAN+LV
Sbjct: 4 IALLIVIAIAFAVLVIYLILLLRKISDTVDESRQTLKILTSVDVNVTLYQTNELLAKANLV 63

50      Query: 66 EDVNGKVATIDPLFVAIALSESLSSDLNSQARHFGQKATNATGNVSKAGKLALVGKVASK 125
        EDVNGKV TIDPLF AIADLS S+SDLN QAR+FG+K +T NV KAG GKVASK
Sbjct: 64 EDVNGKVEITDPLFTAIAADLSVSVDLNRQARYFGKKTRKSTANVGKAGAAAYTFGKVASK 123

55      Query: 126 VFGKKGEK 133
          +F KKG++
Sbjct: 124 LFRKKKGKQ 131
```

An alignment of the GAS and GBS proteins is shown below:

```

Identities = 92/131 (70%), Positives = 116/131 (88%)
60      Query: 1 MIEIAVLIIIAIAFVVVLVLGILFVLKKVSETIEETKQTIKVLTSDVNVTLYQTNEILAKAN 60
          ++ I+++IIA+AFV LV+ ++ VLKKVSETI+E K+TI VLTSVDVNVTL+QTN+ILAKAN
```

-433-

```

Sbjct: 3 LVGISLMIIALAFVALVIFLIVLKKVSETIDEAKKTISVLTSVDNVTLHQTNIDILAKAN 62
Query: 61 VLVDDVNGKVSTIDPLFVAIADLSESVDLNLQARHIGQKASSATSSVTKAGSALAIKGKA 120
      +LV+DVNGKV+TIDPLFVAIADLSES+SDLN QARH GQKA++AT +V+KAG +GK
5 Sbjct: 63 ILVEDVNGKVATIDPLFVAIADLSESLSDLNSQARHFGQKATNATGNVSKACKLALVGKV 122
Query: 121 ASKIFRKKGDK 131
      ASK+F KKG+K
Sbjct: 123 ASKVFGKKGEK 133

```

10

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 342

15 A DNA sequence (GBSx0373) was identified in *S.agalactiae* <SEQ ID 1109> which encodes the amino acid sequence <SEQ ID 1110>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
20 bacterial cytoplasm --- Certainty=0.0462(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

The protein has no significant homology with any sequences in the GENPEPT database.

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 343

30 A DNA sequence (GBSx0374) was identified in *S.agalactiae* <SEQ ID 1111> which encodes the amino acid sequence <SEQ ID 1112>. This protein is predicted to be prolipoprotein diacylglyceryl transferase (lgt). Analysis of this protein sequence reveals the following:

```

Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
      INTEGRAL Likelihood = -8.39 Transmembrane 231 ~ 247 ( 225 ~ 251)
      INTEGRAL Likelihood = -7.64 Transmembrane 89 ~ 105 ( 87 ~ 107)
      INTEGRAL Likelihood = -5.20 Transmembrane 18 ~ 34 ( 13 ~ 36)
      INTEGRAL Likelihood = -1.86 Transmembrane 46 ~ 62 ( 46 ~ 64)

----- Final Results -----
40 bacterial membrane --- Certainty=0.4354(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

45 A related GBS nucleic acid sequence <SEQ ID 9721> which encodes amino acid sequence <SEQ ID 9722> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

>GP: AAC80171 GB:U75480 putative prolipoprotein diacylglycerol
      transferase [Streptococcus mutans] (ver 3)
      Identities = 184/257 (71%), Positives = 226/257 (87%)
50 Query: 2 MINPVAIRLGPFNSIRWYAI CIVSGMILLAVYI LAMKEAPRKNIKSDDILD FILMAFPLSIVG 61
      MINP+AI+LGP +IRWY+ICIV+G++LAVYL ++EAP+KNIKSDD+LD FIL+AFPL+IVG

```

-434-

5 Sbjct: 1 MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIAREAPKKNIKSDDVLDLFILIAFPLAIVG 60
 Query: 62 ARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGGLITGAILLVIFSYRRLINPIDFLDIAA 121
 AR+YYVIF+W YY K+P EI IW+GGIAIYGGL+TGA++L IFSY R+I PIDFLD+AA
 Sbjct: 61 ARLYYYVIFDWYDYYLKNPSEIPVIWGGIAIYGGLLTGALVLFIFSYRMIKPIDFLDVAA 120
 Query: 122 PGVMIAQAIGRWGNFINQEAYGRAVKNLNYVPNFIKNQMYIDGAYRVPTFLYESLWNFLG 181
 PGVM+AQ+IGRWGNF+NQEAYG+ V LNY+P+FI+ QMYIDG YR PTFLYESLWN LG
 Sbjct: 121 PGVMLAQSIGRGWGNFVNQEAYGKVTQLNLYLPDFIRKQMYIDGHYRTPTFLYESLWNLLG 180
 10 Query: 182 FVIIMSIHRHRPRTLKGEGEVACFYLVWYGCGRFIEGMRTDSLVLAGLRVSQWLVSILVII 241
 F+IIM +R RP LK+GEVA FYL+WYG GRF+IEGMRTDSL A LRVSQWLSSV+LV++
 Sbjct: 181 FIIIMILRRRPNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSL MFASLRVSQWLVSLLVVV 240
 15 Query: 242 GIVMIIYRRREQHISYY 258
 G+++++ RRR I YY
 Sbjct: 241 GVIILMVIRRRNHAIPYY 257

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1113> which encodes the amino acid sequence <SEQ ID 1114>. Analysis of this protein sequence reveals the following:

Possible site: 28
 >>> Seems to have an uncleavable N-term signal seq
 25 INTEGRAL Likelihood = -7.01 Transmembrane 229 - 245 (222 - 249)
 INTEGRAL Likelihood = -6.90 Transmembrane 45 - 61 (40 - 68)
 INTEGRAL Likelihood = -4.41 Transmembrane 17 - 33 (11 - 35)
 INTEGRAL Likelihood = -4.14 Transmembrane 87 - 103 (86 - 106)
 INTEGRAL Likelihood = -0.27 Transmembrane 170 - 186 (170 - 186)
 30 ----- Final Results -----
 bacterial membrane --- Certainty=0.3803 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

35 >GP: AAC80171 GB:U75480 putative prolipoprotein diacylglycerol
 transferase [Streptococcus mutans] (ver 3)
 Identities = 176/258 (68%), Positives = 217/258 (83%)
 40 Query: 1 MINPIALKCGPLAIHWYALCILSGLVLAVYLASKEAPKKGISSDAIFDFILIAFPLAIVG 60
 MINPIA+K GPL I WY++CI++GL+LAVYL +EAPKK I SD + DFILIAFPLAIVG
 Sbjct: 1 MINPIAIKLGPLTIRWYSICIVTGLILAVYLTIAREAPKKNIKSDDVLDLFILIAFPLAIVG 60
 Query: 61 ARIYYVIFEWYSYYVHLDEIIAIWNGGIAIYGGLITGALVLLAYCYNKVLNPPIHFLDIAA 120
 AR+YYVIF+W YY+K+ EI IW+GGIAIYGGL+TGA++L IFSY R+I PIDFLD+AA
 45 Sbjct: 61 ARLYYYVIFDWYDYYLKNPSEIPVIWGGIAIYGGLLTGALVLFIFSYRMIKPIDFLDVAA 120
 Query: 121 PSVMVAQAIGRWGNFINQEAYGKAVSQLNLYPSFIQKQMFIEGSYRIPTFLYESLWNLLG 180
 P VM+AQ+IGRWGNF+NQEAYGK V+QLNLYLP FI+KQ+I+G YR PTFLYESLWNLLG
 Sbjct: 121 PGVMLAQSIGRGWGNFVNQEAYGKVTQLNLYLPDFIRKQMYIDGHYRTPTFLYESLWNLLG 180
 50 Query: 181 FVIIMMWRRPKSLLDGEIFAFYLIWYGSGRVIEGMRTDSL MFIRL GIRISQYVSALLIII 240
 F+IIM+ RR+P L +GE+ FYLIWYGSGR VIEGMRTDSL MF +R+SQ++S LL++
 Sbjct: 181 FIIIMILRRRPNLLKEGEVAFFYLIWYGSGRFVIEGMRTDSL MFASLRVSQWLVSLLVVV 240
 55 Query: 241 GLIFVIKRRRQKGISYYQ 258
 G+I ++ RRR I YYQ
 Sbjct: 241 GVIILMVIRRRNHAIPYYQ 258

An alignment of the GAS and GBS proteins is shown below:

60 Identities = 176/257 (68%), Positives = 221/257 (85%)
 Query: 2 MINPVAIRLGPFSIRWYAICIVSGMLLAVYLAMKEAPRKNIKSDDIILDFILMAFPLSIVG 61
 MINP+A++ GP +I WYA+CI+SG++LAVYLA KEAP+K I SD I DFIL+AFPL+IVG
 Sbjct: 1 MINPIALKCGPLAIHWYALCILSGLVLAVYLASKEAPKKGISSDAIFDFILIAFPLAIVG 60

Query: 62 ARIYYVIFEWAYYSKHPVEIIAIWNGGIAIYGLITGAILLVFSYRRLINPIDFLDIAA 121
ARIYYVIFEW+YY KH EIIAIWNGGIAIYGLITGA++L+ + Y +++NPI FLDIAA
Sbjct: 61 ARIYYVIFEWSYVKHLDEIIAIWNGGIAIYGLITGALVLLAYCYNKVLNPPIHFLDIAA 120

5 Query: 122 PGVMIAQAIWRWGNFINQEAYGRAVKNLNYVPNFIKNQMYIDGAYRVPTFLYESLWNFLG 181
P VM+AQAIGRWRGNFINQEAYG+AV LNY+P+FI+ QM+I+G+YR+PTFLYESLWN LG
Sbjct: 121 PSVMVAQAIWRWGNFINQEAYGKAVSQLNYLPSFIQKQMFIEGSYRIPPTFLYESLWNLLG 180

10 Query: 182 FVIIMSIHRPRTLKQGEVACFYLVWYGCGRFIIEGMRTDSL YLAGLRLVSQWLNSVILVII 241
FVIIM R +P++L GE+ FYL+WYG GR +IEGMRTDSL G+R+SQ++S +L+II
Sbjct: 181 FVIIMMWRRKPKSLLDGEIFAFYLIWYGSGLVIEGMRTDSLMLGIRISQYVSALLIII 240

15 Query: 242 GIVMIIYRRREQHISYY 258
G++ +I RRR++ ISYY
Sbjct: 241 GLIFVIKRRRQKGISYY 257

A related GBS gene <SEQ ID 8557> and protein <SEQ ID 8558> were also identified. Analysis of this protein sequence reveals the following:

20 Lipop: Possible site: -1 Crend: 0
McG: Discrim Score: 2.45
GvH: Signal Score (-7.5): -2.9
Possible site: 39
>>> Seems to have an uncleavable N-term signal seq
25 ALOM program count: 3 value: -8.39 threshold: 0.0
INTEGRAL Likelihood = -8.39 Transmembrane 209 - 225 (203 - 229)
INTEGRAL Likelihood = -7.64 Transmembrane 67 - 83 (65 - 85)
INTEGRAL Likelihood = -1.86 Transmembrane 24 - 40 (24 - 42)
30 PERIPHERAL Likelihood = 0.79 92
modified ALOM score: 2.18
*** Reasoning Step: 3
----- Final Results -----
35 bacterial membrane --- Certainty=0.4354 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

40 ORF01400 (238 - 1008 of 1308)
SP|P72482|LGT_STRMU(1 - 257 of 259) PROLIPOPROTEIN DIACYLGLYCERYL TRANSFERASE (EC 2.4.99.-). GP|4583534|gb|AAC80171.3|U75480 putative prolipoprotein diacylglycerol transferase {Streptococcus mutans} PIR|T11569|T11569 prolipoprotein diacylglyceryl transferase (EC 2.4.99.-) - Streptococcus mutans
45 %Match = 46.9
%Identity = 71.6 %Similarity = 89.5
Matches = 184 Mismatches = 27 Conservative Sub.s = 46
50 198 228 258 288 318 348 378 408
WGLMLPRLRIV*HI*LVRTRSMMINPVPAIRLGPFSIRWYAJICIVSGMLLAVYLAMKEAPRKNIKSDDILDFIGMAFPLS
|||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
MINPIAIKLGPLTIRWYSICIVTGLILAVLYTIREAPKKNIKSDDVLDFIGLIAFPLA
10 20 30 40 50
55 438 468 498 528 558 588 618 648
IVGARLYYYVIFDWYDYLKNPSEIPVIWHGGIAIYGLLTGAIQLFIFSYRMIKPIDFLDVAAPGVMLAQSIGRWGNFVN
|||||:||||:||| :||| :||| ||:|||||:|||:||| :||| :||| :||| :||| :||| :||| :||| :||| :|||:
IVGARLYYYVIFDWYDYLKNPSEIPVIWHGGIAIYGLLTGAIQLFIFSYRMIKPIDFLDVAAPGVMLAQSIGRWGNFVN
70 80 90 100 110 120 130
60 678 708 738 768 798 828 858 888
QEAYGRAVKNLNYVPNFIKNQMYIDGAYRVPTFLYESLWNFLGFVIMSIRHRPRTLKQGEVACFYLVWYGCGRFIIEGM
|||||: | | ||:|||: |||:||| |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QEAYGKTVTQLNLPDFIRKQMYIDGHYRTPTFLYESLWNLLGFIIIMLRRRNPNLLKEGEVAFFYLWYGSGRFVIEGM
65 150 160 170 180 190 200 210

918 948 978 1008 1038 1068 1098 1128
 RTDSDLYLAGLDRVSQLSVILVIIGIVMIIYRRREQHISYY*TEEVL**KLLY*LLPLRLLF*F*EYFSF*KKYQKRLRKP
 ||||| :| |||||||||:||::|:::| ||| :| |||
 5 RTDSDLMFASLRVSQLSVLLVVGVILMVIRRNRHAIPYYQC
 230 240 250

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 344

- 10 A DNA sequence (GBSx0375) was identified in *S.agalactiae* <SEQ ID 1115> which encodes the amino acid sequence <SEQ ID 1116>. Analysis of this protein sequence reveals the following:

Possible site: 31
 >>> Seems to have no N-terminal signal sequence

15 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2817 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

- 20 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA77782 GB:AB027460 Hpr kinase [Streptococcus bovis]
 Identities = 264/309 (85%), Positives = 292/309 (94%)

25 Query: 1 MAVTVQMLVDRLLKLNVIYGDHEHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60
 M+VTV+MLVD++KL+VIYGD+ LLSK ITT+DISRPGLEMTGYFDYY+PERLQL+GMKEW
 Sbjct: 1 MSVTVKMLVDVKLDVIYGDDLLSKEITTS DISRPGLEMTGYFDYYSPERLQLLGMEW 60

 Query: 61 SYLMAMTGHNRYQVLREMFQKETPAIVVARDLEIPEEMYFAAKDTGIAILQSKAPTSRLS 120
 SYL MT HNR VLREM + ETPAI+VAR+L IPEEM AAK+ GIAILQS PTSRLS
 30 Sbjct: 61 SYLTAKMTSHNRRHVLRREMIKPTPAIIIVARNLAIPEEMISAKEKGIAILQSHVPTSRLS 120

 Query: 121 GEVSWYLDSCLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGRHLVADDRVD 180
 GE+SWYLDSCLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGRHLVADDRVD
 Sbjct: 121 GEMSWYLDSCLAERTSVHGVLMDIYGMGVLIQGDSGIGKSETGLELVKRGRHLVADDRVD 180

 Query: 181 VYAKDEETLWGEPAEILRHLLERGVGIIDIMSLYGASAVKDSSQVQLAIYLENFETGKV 240
 V+AKDEETLWGEPAEILRHLLERGVGIID+MSLYGASAVKDSSQVQLAIYLEN+E+GKV
 Sbjct: 181 VFAKDEETLWGEPAEILRHLLERGVGIIDVMSLYGASAVKDSSQVQLAIYLENEYSGKV 240

 40 Query: 241 FDRLGNGNEEIELSGVKVPRIRIPVKTGRNVSVVIEAAAMNHRAKQMGFDATQTFEDRLT 300
 FDRLGNGNEE+ELSGVK+PR+RIPV+TGRN+SVVIEAAAMN+RAKQMGFDAT+TFE+RLT
 Sbjct: 241 FDRLGNGNEEIELSGVKIPRLRIPVQTGRNMSVVIEAAAMNYRAKQMGFDATKTFEERLT 300

 45 Query: 301 HLISQNEVN 309
 LI++NE N
 Sbjct: 301 QLITKNEGN 309

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1117> which encodes the amino acid sequence <SEQ ID 1118>. Analysis of this protein sequence reveals the following:

50 Possible site: 13
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2391 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 255/309 (82%), Positives = 288/309 (92%)

-437-

Query: 1 MAVTVQMLVDRLKLNVIYGDEHLLSKRITTADISRPGLEMTGYFDYYAPERLQLVGMKEW 60
M VTV+MLV ++KL+V+Y ++LLSK ITT+DISRPGLEMTGYFDYYAPERLQL GMKEW
Sbjct: 32 MTVTVKMLVQKVQLDVYATDNLLSKEITTSDISRPGLEMTGYFDYYAPERLQLFGMKEW 91

5 Query: 61 SYLMAMTGHNRQVILREMFKETPAIVVARDLEIPEEMYEAAKDTGIAILQSKAPTSRLS 120
SYL MT HNRY VI+EMF+K+TPA+VV+R+L IP+EM +AAK+ GI++L S+ TSRL+
Sbjct: 92 SYLTQMTSHNRYSVILKEMFKKDTPAVVVSRNLAIPKEMVQAKEEGISLLSSRVSTSRLA 151

10 Query: 121 GEVSWYLDSCLAERTSVHGVLMIDIYGMGVLIQGDSGIGKSETGLELVKRGRHLVADDRVD 180
GE+S++LD+ LAERTSVHGVLMIDIYGMGVLIQGDSGIGKSETGLELVKRGRHLVADDRVD
Sbjct: 152 GEMSYFLDASLAERTSVHGVLMIDIYGMGVLIQGDSGIGKSETGLELVKRGRHLVADDRVD 211

15 Query: 181 VYAKDEETLWGEPAEILRHLLEIRGVGIIDIMSLYGASAVKDSSQVQLAIYLENFTGKV 240
VYAKDEETLWGEPAEILRHLLEIRGVGIID+MSLYGASAVKDSSQVQLAIYLENFE GKVK
Sbjct: 212 VYAKDEETLWGEPAEILRHLLEIRGVGIIDVMSLYGASAVKDSSQVQLAIYLENFEAGKV 271

20 Query: 241 FDRLGNGNEEIELSGVKVPRIRIPVKTGRNVSVVIEAAAMNHRAKQMGFDATQTFEDRLT 300
FDRLGNGNEEI SGV+ +PRIRIPVKTGRNVSVVIEAAAMNHRAK+MGFDAT+TFEDRLT
Sbjct: 272 FDRLGNGNEEITFSQVRIPRIRIPVKTGRNVSVVIEAAAMNHRAKEMGFDATKTFEDRLT 331

25 Query: 301 HLISQNEVN 309
LI++NEV+
Sbjct: 332 QLITKNEVS 340

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 345

A DNA sequence (GBSx0376) was identified in *S.agalactiae* <SEQ ID 1119> which encodes the amino acid sequence <SEQ ID 1120>. Analysis of this protein sequence reveals the following:

Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35 bacterial cytoplasm --- Certainty=0.1836 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9719> which encodes amino acid sequence <SEQ ID 9720> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 346

A DNA sequence (GBSx0377) was identified in *S.agalactiae* <SEQ ID 1121> which encodes the amino acid sequence <SEQ ID 1122>. Analysis of this protein sequence reveals the following:

Possible site: 37
>>> Seems to have an uncleavable N-term signal seq
50 INTEGRAL Likelihood = -4.88 Transmembrane 35 - 51 (31 - 59)

----- Final Results -----
bacterial membrane --- Certainty=0.2954 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

-438-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

```

5      >GP: AAC67275 GB: AF017113 YvlC [Bacillus subtilis]
      Identities = 21/63 (33%), Positives = 36/63 (56%), Gaps = 2/63 (3%)
      Query: 3 SSFYKQRKGKLVCGVVAGLADKYNDLALSRLVIALILYFTKF--GLLLYILLAVFLPYK 60
              + Y+ K K + GV+ GLA+ +NWD +L RV+ ++ T LL+YI+ +P +
      Sbjct: 2 NKLYRSEKNKKIAGVIGGLAEYFNWDASLLRVITVILAIMTSVLPVLLIYIIWIFIVPSE 61
      Query: 61 EDI 63
              D+
      Sbjct: 62 RDM 64
  
```

- 15 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1123> which encodes the amino acid sequence <SEQ ID 1124>. Analysis of this protein sequence reveals the following:

```

20      Possible site: 32
      >>> Seems to have an uncleavable N-term signal seq
          INTEGRAL Likelihood = -5.26 Transmembrane 39 - 55 ( 31 - 61)
      ----- Final Results -----
      bacterial membrane --- Certainty=0.3102 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
  
```

25 An alignment of the GAS and GBS proteins is shown below:

```

25      Identities = 60/90 (66%), Positives = 77/90 (84%), Gaps = 3/90 (3%)
      Query: 1 MKSSFYKQRKGKLVCGVVAGLADKYNDLALSRLVIALILYFTKFGLLLYILLAVFLPYK 60
              +++ FYKQRK +LV GV+AGLADKY WDLAL+RVL AL++Y T FG+LLYILLA+FLPYK
      Sbjct: 1 VETKFYKQRKNRLVAGVIAGLADKYGWDLALARVLAALLIYGTGFGVLLYILLAIFLPYK 60
      Query: 61 EDIIETR-RQGP RRKD AEPV--DDDGF FW 87
              ED++E R +GPRRKDA+ + ++DGFWF
      Sbjct: 61 EDLLEERYGRGP RRKD ADVLNEEDGFWF 90
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 347

- 40 A DNA sequence (GBSx0378) was identified in *S.agalactiae* <SEQ ID 1125> which encodes the amino acid sequence <SEQ ID 1126>. Analysis of this protein sequence reveals the following:

```

45      Possible site: 19
      >>> Seems to have no N-terminal signal sequence
      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3577 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

- 50 A related GBS nucleic acid sequence <SEQ ID 9717> which encodes amino acid sequence <SEQ ID 9718> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```

55      >GP: BAB04250 GB: AP001508 unknown conserved protein [Bacillus halodurans]
      Identities = 379/729 (51%), Positives = 515/729 (69%), Gaps = 25/729 (3%)
      Query: 29 ENLNITQIAIDLGIKASQIEKVLELTDEGN TIPFIARYRKEMTGNLDEVQIKSIIDL DK 88
  
```

-439-

E I +A +L +K + I++V++L EGNT+PFIARYRKE+TG +DEV+I+ + +
 Sbjct: 8 EEHTIKTLAKELSLKPNEYIKQVIQLLHEGNTVPFIARYRKELTGGMDEVKIREVSEKWTY 67

5 Query: 89 MTALSDRKTTVLAKIEEQGKLTQELKKAIEEATKLADVEEELYLPYKEKRRRTKATIAREAG 148
 L +RK V+ +EEQGKLT E KK +E+A KL +VE+LY PYK+KRRT+AT+A+E G
 Sbjct: 68 ANQLHERKEEVIRLVEEQGKLTDEWKKTVEQAQKLQEVEDLYRPyKQKRRTRATVAKEKG 127

10 Query: 149 LFPLARLI--LQNNDNLEEEAQNYLTDGFETTT--KALSGAVDILIEAFSEDNKLRSWTY 204
 L PLA + L + EA+ YL+ E T L GA DI+ E ++D LR
 Sbjct: 128 LEPLAEWLFSLPRDGDPLQEAEVYLSVEHELTKVEDVLQGAQDIIAEWIADDADLRKRIR 187

15 Query: 205 NEIWNYSITAVVKDESDEKQVFKYYDFSEKISKLHGYQVILANRGEKMGVLKVNFEH 264
 + + S+ A VK E LDEK V+++YYD+ E + L ++ LALNRGEK VL+V
 Sbjct: 188 SLGFKEGSVIAKVKKEELDEKGVYEMYYDYEPPVRTLVPHRTLALNRGEKEDVLRVTIRF 247

20 Query: 265 NLEKMFRF---FAVRFKETS-QYIDDLIVQTVKKKIVPAMERRIRTELSEGAEDGAISL 319
 ++++ F RF + Y+ I K+ I P++ER IR EL+E AE+ AI +
 Sbjct: 248 PVDRIIEMSEKTFIRRGSPAVPYVKAIAEDGYKRLIEPSIERIRHELTEKAEEQAIHI 307

25 Query: 320 FSENLRNLLLVSPLKGKMWLGDPAPRTGAKLAVVDQTKLMTTQVIYPVPPANQAKIEQ 379
 F+ENLR+LLL P+KGK+VLG DPA+RTG KLA+VD+TGA+ P QVIYP PP N+ +
 Sbjct: 308 FAENLRSLLOPPIKGKVVLGLDPAYRTGCKLAIIVDETGKVLDIQVIYPTPPKNE--VAA 365

30 Query: 380 SKIELAKLIKEFNIEIIAIGNGTASRESEAFVAEVLQDFPD-VSYVIVINESGASVYSASE 438
 +K + KLI ++ +E+IAIGNGTASRESE F+A++++D P + Y+IVINE+GASVYSASE
 Sbjct: 366 AKKIVKKLIADYGVEMIAIGNGTASRESEQFIADLIKDLQTIYYLIVNEAGASVYSASE 425

35 Query: 439 LARHEFPDLTVEKRSASIARRLQDPLAELVKIDPKSIGVGQYQHDVSQKKLAENLDFVV 498
 + R EFPDL VE+RSA+SIARRLQDPLAELVKIDPKS+VGQYQHDVSQK+L E+L FVV
 Sbjct: 426 IGREEFPDLQVEERSAVSIARRLQDPLAELVKIDPKSVGQYQHDVSQKRLNESLTFVV 485

40 Query: 499 ETVVNQVGVNNTASPALLAHVSGLNKTISENIVKYREENGQIKSRAEIKKVPRLGAKAF 558
 ETVVNQVGVNNTASP+LL +V+GL+KT+++NIVK REE G +RA++K +PRLGAK +
 Sbjct: 486 ETVVNQVGVNNTASPSSLLQYVAGLSKTVAKNIVKKREEAGRFTARAQLDIPRLGAKTY 545

45 Query: 559 EQAACFLRIENAKNFLDTIGVHPESYEAVKKLLDQLTIKELD---DLAKEKLQNLDDLIAT 615
 EQ GFLRI + N LD T +HPESY+ KLL ++ D + K+KLQ LD+ A
 Sbjct: 546 EQCIGFLRIMGDNLNLDATAIHPESYKVTDKLSEVGATAADVGIEDLKKLQALDVSAM 605

50 Query: 616 AESIGVGQETLKDIIEDLILKPGDRDLRDDFEAPVLRHDVLDVSDLKVGQELQGTVRNVDF 675
 A ++ VG TLKD+I+ L++P RD RD+ P+L+ DVL + DL G ELQGTVRNVDF
 Sbjct: 606 AATLDVGVPVTLKDMIDALIRPTRDPRDEVAKPLLQDVLQLEDLLPGMELQGTVRNVDF 665

Query: 676 GAFVDIGVHEDGLIHQSRLIKRKDRKKTRKMPPLQHPSKYLGVGDIVTVWWVEVDAERSR 735
 G FVDIGV +DGL+H S+L R ++HP + ++VG+IVTVWW +VD ++ R
 Sbjct: 666 GVFVDIGVKQDGLVHISKLANRY-----IKHPLEVVTVCIEIVTVWWVEDVDIKKGR 715

Query: 736 IGLSLIKPD 744
 I L++++P+
 Sbjct: 716 IAL/TMLRPE 724

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1127> which encodes the amino acid sequence <SEQ ID 1128>. Analysis of this protein sequence reveals the following:

Possible site: 25

55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2207(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 532/716 (74%), Positives = 619/716 (86%), Gaps = 10/716 (1%)

65 Query: 28 MENLNITQIAIDLGIKASQIEKVLLELTDEGNTIPFIARYRKEMTGNLDEVQIKSIIDLK 87

-440-

MEN N IA L + QIE+VL LT +GNTIPFIARYRKE+TGNLDEV IKSIID+DK
 Sbjct: 1 MENNNNHNIAEALSVSLHQIEQVLAQGNTIPFIARYRKEVTGNLDEVVIKSIIDMDK 60

5 Query: 88 SMTALSDRKTTVLAKIEEQGKLTQELKKAIEEATKLADVEELYLPYKEKRRTKATIAREA 147
 S+T L++RK T+LAKIEEQGKLT +L+ +IE KLAD+EELYLPYKEKRRTKATIAREA
 Sbjct: 61 SLTTLNERKATILAKIEEQGKLTDQLRTSIEATEKLADLEELYLPYKEKRRTKATIAREA 120

10 Query: 148 GLFPLARLILQNQDNLEEEAQNYLTDFETTTKALSGAVDILIEAFSEDNKLRSWTYNEI 207
 GLFPLARLILQNQDNLE A+ ++T+GF + +AL+GAVDIL+EA SED KLRSWTYNEI
 Sbjct: 121 GLFPLARLILQNQDNLEAAEPFVTEGFASPOREALAGAVDILVEAMSEDAKLRSWTYNEI 180

15 Query: 208 WNYSSITAVVKDESLDEKQVFKIVYDFSEKISKLHGYQVLALNRGEKGMLVKNFEHNLE 267
 W YS + + KDE LDEK+VF+IYDFS++S + GY+ LALNRGEK+G+LKV+FEHNLE
 Sbjct: 181 WQYSRLVSTLKDQLDEKKVFQIYYDFSDQVSNMQGYRTLALNRGEKLGILKVSFEHNLE 240

20 Query: 268 KMFRFFAVRFKETSQYIDDLIVQTVKKKIVPAMERRIRTELSEGAEDGAIISLFSENLRNL 327
 KM RFF+VRFKET+ YI++I QT+KKKIVPAMERR+R+ELS+ AEDGAI LFSENLR+L
 Sbjct: 241 KMQRFFSVRFKETNPYIEEVINQTIKKKIVPAMERRVRSELSAAEDGAIHLFSENLRHL 300

25 Query: 328 LLVSPKGKVMVLGFDPAFRTGAKLAIVDQTKLTTQVIYPVAPASQTKIQAAKETLTQ 387
 LLVSPKGKVMVLGFDPAFRTGAKLA+VDQTKL+TTQVIYPV PA+Q KI+ +K L +L
 Sbjct: 301 LLVSPKGKVMVLGFDPAFRTGAKLAIVDQTKLTTQVIYPVAPASQTKIQAAKETLTQ 360

30 Query: 388 IKEYNIEIIAIGNGTASRESEAFVAEVLQDFPDVSYVIVNESGAVSVSASELARHEFPDL 447
 I+ + I+IIAIGNGTASRESEAFVA+VL+DFP+ SYVIVNESGAVSVSASELARHEFPDL
 Sbjct: 361 IETYQIDIICAIIGNGTASRESEAFVADVLKDFPNTSYVIVNESGAVSVSASELARHEFPDL 420

35 Query: 448 TVEKRSASIARRLQDPLAELVKIDPKSIGVGQQYQHDVSQKQLAEIDLDFVVETVVNVQGV 507
 TVEKRSASIARRLQDPLAELVKIDPKSIGVGQQYQHDVSQKQL+ENL FVV+TVVNQGV
 Sbjct: 421 TVEKRSASIARRLQDPLAELVKIDPKSIGVGQQYQHDVSQKQLSENLGVVDTVVNVQGV 480

40 Query: 508 NVNTASPALLAHVSGLNKTISENIVKYREENGQIKSRAEIKVPRLGAKAFEQAGFLRI 567
 NVNTASP+LLAHVSGLNKTISENIVKYREENG + SRA+IKVPRLGAKAFEQAGFLRI
 Sbjct: 481 NVNTASPSSLLAHVSGLNKTISENIVKYREENGALTSRADIKVPRLGAKAFEQAGFLRI 540

45 Query: 568 PNAKNFLDNTGVHPESYEAVKLLDQLTIKEELDDLAKEKLQNLDDLIATAESIGVGQETLK 627
 P AKN LDNTGVHPESY AVK+L L I++LDD AK L + + AE++ +GOETLK
 Sbjct: 541 PGAKNILDNTGVHPESYPAVKELFKVLGIQDLDAAKATLAAQVQPQMAETLAIGQETLK 600

50 Query: 628 DIIEDLLKPGRDLRDDFEAPVLRHDVLDVSDLKVQQLQGTVRNVVDFGAFVDIGVHEDG 687
 DII DLLKPGRDLRDDFEAP+LR D+LD+ DL++GQ+L+GTVRNVVDFGAFVDIGVHEDG
 Sbjct: 601 DIIADLLKPGRDLRDDFEAPILRQDILDLKDLEIGQKLEGTVRNVVDFGAFVDIGVHEDG 660

55 Query: 688 LIHQSRSLIKRKDRKTRKMPPLQHPSKYLQGDIVTVWWVVEDAERSRIGLSLIK 743
 LIH S + K + HPS+ +SVGD+VTWV + +D +R ++ LSL+ P
 Sbjct: 661 LIHISEMSKTF-----VNHPSQVSVGDLTVWVSKIDLDRHKVNLSLLPP 706

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

50 Example 348

A DNA sequence (GBSx0379) was identified in *S.agalactiae* <SEQ ID 1129> which encodes the amino acid sequence <SEQ ID 1130>. This protein is predicted to be N5,N10-methylenetetrahydromethanopterin reductase homolog. Analysis of this protein sequence reveals the following:

Possible site: 60
 55 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4864 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

60 The protein has homology with the following sequences in the GENPEPT database:

-441-

>GP:AAB94650 GB:U96107 N5,N10-methylenetetrahydromethanopterin
reductase homolog [Staphylococcus carnosus]
Identities = 164/300 (54%), Positives = 217/300 (71%), Gaps = 1/300 (0%)

5 Query: 45 VYGIGEHHRDFAVSAPEIVLAAAGAVRTNNIRLSSAVTILSSNDPIRVYQQFSTIDALSN 104
+YG+GEHHR D+AVS P VLAA A T I+LSSAVT+LSS+DP+ VY++F+T+DA+SN
Sbjct: 1 MYGLGEHHRSDYAVSDPVTVLAAAASLTQRIKLSSAVTVLSSDDPVCVYERFATLDAVSN 60

10 Query: 105 GRAETMAGRGSFIESFPLFGYDLADYDDLNFNEKMDMILLAINSATNLDWKGHLTQTVNERP 164
GRAEIM GRGSFIESFPLFGYDL DYD LF EK+++L IN + W+G + +
Sbjct: 61 GRAETMVGGRGSFIESFPLFGYDLDDYDRLFVEKLELLKEINQHEVVTWEGTMRPAIKGLG 120

15 Query: 165 IYPRALQRQLPIWVATGGNVDSTIRIAEQGLPIVYATIGGNPKAFRQLVHIYKEVGSRNG 224
+YPRA+Q ++PIW+ATGG +S+IR AE GLPI YA IGGNPK F++ + IY+ V G
Sbjct: 121 VYPRAVQDEIPIWLATGGTPESSIRAAEFGLPITYAIIGGNPKRFKRNIAIYRAVAESRG 180

20 Query: 225 HKPEQLKVAAHSGWGWEEDNQTAIDRYFFPTKQTVDNIAKGRPHWSEMTKEQYLRSVGPE 284
+ + VA HSWG+I + ++ A ++ PTK + IAK R +W T+ + R + E
Sbjct: 181 YDLADMPVAVHSGYIADTDEQAQREFYEPTKVHHEIIAKER-NWPPYTEAHFQREISDE 239

Query: 285 GAIFVGSPPEVVAHKIIGLVEALELDRFMLHLPGSMPHKDVLNAIKLYGKEVAPIVRKF 344
GA+FVGSPPE VA K+I ++E L L+RFMLH+PGSMPH+ ++ AIKLYGK V PI+ YF
Sbjct: 240 GAMFVGSPETVARKMIKVIIEELGLNRFMLHIPVGSMRPERIMKAIKLYGKRVKPIIEDYF 299

25 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 349

A DNA sequence (GBSx0380) was identified in *S.agalactiae* <SEQ ID 1131> which encodes the amino acid sequence <SEQ ID 1132>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
35           bacterial cytoplasm --- Certainty=0.1310(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9715> which encodes amino acid sequence <SEQ ID 9716> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1133> which encodes the amino acid sequence <SEQ ID 1134>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
45           >>> Seems to have no N-terminal signal sequence

----- Final Results -----
          bacterial cytoplasm --- Certainty=0.0915(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
50           bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

Identities = 20/40 (50%), Positives = 27/40 (67%), Gaps = 3/40 (7%)

55 Query: 4 MAITHKRQDDLESFASFAKVP---KPKKVDSDSKPEQKD 40
MAITHK+ D+LE M A FA +P KP +V++D K K+

Sbjct: 1 MAITHKKNDELEKMLAGFASIPSFDKPLEVNTDGKLATKE 40

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

5 Example 350

A DNA sequence (GBSx0381) was identified in *S.agalactiae* <SEQ ID 1135> which encodes the amino acid sequence <SEQ ID 1136>. Analysis of this protein sequence reveals the following:

Possible site: 56

>>> Seems to have no N-terminal signal sequence

10

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1453 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

15

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

20 Example 351

A DNA sequence (GBSx0382) was identified in *S.agalactiae* <SEQ ID 1137> which encodes the amino acid sequence <SEQ ID 1138>. Analysis of this protein sequence reveals the following:

Possible site: 37

>>> Seems to have an uncleavable N-term signal seq

25

INTEGRAL	Likelihood = -11.15	Transmembrane	216 - 232 (210 - 240)
INTEGRAL	Likelihood = -9.18	Transmembrane	15 - 31 (10 - 39)
INTEGRAL	Likelihood = -9.02	Transmembrane	283 - 299 (276 - 299)
INTEGRAL	Likelihood = -8.76	Transmembrane	128 - 144 (119 - 150)
INTEGRAL	Likelihood = -4.62	Transmembrane	243 - 259 (237 - 265)
INTEGRAL	Likelihood = -2.44	Transmembrane	65 - 81 (65 - 81)
INTEGRAL	Likelihood = -2.44	Transmembrane	94 - 110 (93 - 111)

30

----- Final Results -----

bacterial membrane --- Certainty=0.5458 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

35

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]

40

Identities = 141/283 (49%), Positives = 198/283 (69%), Gaps = 3/283 (1%)

Query: 10 SVLQWFAIFISIIIEALPFVLLGTTILSGIIEVFITPDIVNKFLPKNKFLRVLFGTFVGTV 69
 S LQ +IFISI+IEA+PF+L+G ILSGII++F++ +++ + +PKN+FL VLFG G +

45

Sbjct: 6 SFLQLNSIFISIISILIEAIPFILIGVILSGIIQMVFSEEMIARIMPKNRFLAVLFGALAGVL 65

Query: 70 FPSCECGIPIPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSIRFLIRFVG 129
 FP+CECGIPIPI R L K VP + V F+ TAPIINPIVLF+TY AFGN + R

Sbjct: 66 FPACECGIPIPITRRLLLKGVPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125

50

Query: 130 ATIVAIALGVMLAFLVDDNLKEDAKPHTFHHDYSDKKWYQKIFLALAHAIIDEFFDTGRYL 189
 A V++ +GV+L++ DN L + +P H H + OK+ L HAIDEFF G+YL

Sbjct: 126 ALAVSIIIGVILSYQFKDNQLLPDEPGHHHHGTL-LQKLGGLRHAIDEFFSVGKYL 184

Query: 190 VFGTLLIASAMQIYLPTRVLTTIGHSPITAILVMMILLAFLSLCSEADAFIGASLLSTFGI 249

-443-

+ G IA+AMQ Y+ T L IG + +++ LVMM LAF+LSLCSE DAFI +S STF +
 Sbjct: 185 IIGAFIAAAAMQTYVKTSTLLAIGQNDVSSLVMMGLAFVSLCSEVDAFIASSFSSTFSL 244

5 Query: 250 APVMAFLLIGPMIDIKNLMMMVNSFKTRFIVQFISVSSLIIII 292
 ++AFL+ G M+DIKNL+MM+ +FK RF+ F+ ++ +++I+
 Sbjct: 245 GSLIAFLVFGAMVDIKNLLMMMLAAFKKRFL--FLLITYIVVIV 285

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1139> which encodes the amino acid sequence <SEQ ID 1140>. Analysis of this protein sequence reveals the following:

10 Possible site: 25
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -9.92 Transmembrane 216 - 232 (211 - 237)
 INTEGRAL Likelihood = -9.45 Transmembrane 283 - 299 (276 - 299)
 INTEGRAL Likelihood = -8.76 Transmembrane 128 - 144 (119 - 150)
 15 INTEGRAL Likelihood = -7.80 Transmembrane 15 - 31 (10 - 39)
 INTEGRAL Likelihood = -5.47 Transmembrane 243 - 259 (237 - 265)
 INTEGRAL Likelihood = -2.44 Transmembrane 65 - 81 (65 - 81)
 INTEGRAL Likelihood = -2.44 Transmembrane 94 - 110 (93 - 111)

20 ----- Final Results -----
 bacterial membrane --- Certainty=0.4970(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

25 The protein has homology with the following sequences in the databases:

>GP:CAB12119 GB:Z99105 ycgR [Bacillus subtilis]
 Identities = 143/288 (49%), Positives = 196/288 (67%), Gaps = 1/288 (0%)

30 Query: 10 SVLQWFAIFMSIIIEALPFVLLGTILSGCIEVFVTPELVQKYLPKQLRILFGTFVGTV 69
 S LQ +IF+SI+IEA+PF+L+G ILSG I++FV+ E++ + +PK + L +LFG G +
 Sbjct: 6 SFLQLNSIFISILIEAIPFILIGVILSGIIQMVSSEMIARIMPKNRFLAVLFGALAGVL 65

Query: 70 FPSCECGIIPPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSLRFLILRLVG 129
 FP+CECGIIPPI R L K VP + V F+ TAPIINPIVLF+TY AFGN + R
 35 Sbjct: 66 FPACECGIIPITRRLLLKGVPPLHAGVAFMLTAPIINPIVLFSTYIAFGNRWSVVFYRGGL 125

Query: 130 AALVAITLGVMLAFTIVDDNILKDNAQPVHFHDYSHESLPKRITYLALVHAIDEFFDTGRYL 189
 A V++ +GV+L++ DN L +P H H + H +L +++ L HAIDEFF G+YL
 40 Sbjct: 126 ALAVSLIITGVILSYQFKDNQLLKPDEPGH-HHHHHGTLQKLGGTLRHAIDEFFSVGKYL 184

Query: 190 VFGTLIASAMQIYVPTVRVLTIGHNPLTAILIMMLMAFILSLCSEADAFIGASLLSTFGV 249
 + G IA+AMQ YV T L IG N +++ L+MM +AF+LSLCSE DAFI +S STF +
 Sbjct: 185 IIGAFIAAAAMQTYVKTSTLLAIGQNDVSSLVMMGLAFVSLCSEVDAFIASSFSSTFSL 244

45 Query: 250 APVMAFLLIGPMIDIKNLMMMVKAFKGRFIVQFIGVSVLMAVYCLLV 297
 ++AFL+ G MVDIKNL+MM+ AFK RF+ I V+++ LLV
 Sbjct: 245 GSLIAFLVFGAMVDIKNLLMMMLAAFKKRFLVLLITYIVVIVLAGSLV 292

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 248/300 (82%), Positives = 278/300 (92%)

Query: 1 MDIFNQLPDSVLQWFAIFFISIIIEALPFVLLGTILSGIIIEVFITPDIVNKFLPKNKFLRV 60
 M +F+ LP SVLQWFAIF+SIIIEALPFVLLGTILSG IEVF+TP++V K+LPK K LR+
 55 Sbjct: 1 MSLFSNLPPSVLQWFAIFMSIIIEALPFVLLGTILSGCIEVFVTPELVQKYLPKQLR 60

Query: 61 LFGTFVGFVFPSCECGIIPPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSI 120
 LFGTFVGFVFPSCECGIIPPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNS+
 Sbjct: 61 LFGTFVGFVFPSCECGIIPPIINRFLEKKVPSYTAVPFLATAPIINPIVLFATYSAFGNSI 120

60 Query: 121 RFLILRFVGATIVAIAGVMLAFLVDDNILKEDAKPTHFHDYSDKKWYQKIFLALAHAI 180
 RFLILR VGA +VAI LGVMLAF+VDDNILK+A+P HFHDYS + ++I+LAL HAID
 Sbjct: 121 RFLILRLVGAALVAVITLGVMLAFTIVDDNILKDNAQPVHFHDYSHESLPKRITYLALVHAID 180

Query: 181 EFFDTGRYLVFGTLIASAMQIYLPTRVLTTIGHSPITAILVMMLLAFILSLCSEADAFIG 240

-444-

```

EFFDTGRYLVFGTLIASAMQIY+PTRVLTTIGH+P+TAIL+MML+AFTLSLCSEADAFIG
Sbjct: 181 EFFDTGRYLVFGTLIASAMQIYVPTRVLTTIGHNPLTAILIMMLMAFILSLCSEADAFIG 240

Query: 241 ASLLSTFGIAPVMAFLLIGPMIDIKNLMMMVNSFKTRFIVQFISVSSLIIIIYCLFVGVI 300
         ASLLSTFG+APV+AFLLIGPM+DIKNLMMMV +FK RFIVQFI VS L+I +YCL VGV+
Sbjct: 241 ASLLSTFGVAPVLAFLLIGPMVDIKNLMMMVAKFKGRFIVQFIGVSVLMIAVYCLLVGVL 300

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

10 Example 352

A DNA sequence (GBSx0383) was identified in *S.agalactiae* <SEQ ID 1141> which encodes the amino acid sequence <SEQ ID 1142>. Analysis of this protein sequence reveals the following:

```

Possible site: 13
>>> Seems to have no N-terminal signal sequence
15
----- Final Results -----
bacterial cytoplasm --- Certainty=0.4703 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
20

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

25 Example 353

A DNA sequence (GBSx0384) was identified in *S.agalactiae* <SEQ ID 1143> which encodes the amino acid sequence <SEQ ID 1144>. Analysis of this protein sequence reveals the following:

```

Possible site: 50
>>> Seems to have an uncleavable N-term signal seq
30
    INTEGRAL Likelihood = -8.44 Transmembrane 45 - 61 ( 39 - 65)
    INTEGRAL Likelihood = -8.12 Transmembrane 83 - 99 ( 77 - 101)
    INTEGRAL Likelihood = -0.00 Transmembrane 2 - 18 ( 1 - 19)

----- Final Results -----
35
    bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

A related GBS nucleic acid sequence <SEQ ID 8559> which encodes amino acid sequence <SEQ ID 8560> was also identified. Analysis of this protein sequence reveals the following:

```

Lipop Possible site: -1 Crend: 2
SRCFLG: 0
McG: Length of UR: 8
      Peak Value of UR: 2.23
45
      Net Charge of CR: 1
McG: Discrim Score: 0.46
GvH: Signal Score (-7.5): -3.54
      Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
50
Amino Acid Composition: calculated from 1
ALOM program count: 2 value: -8.44 threshold: 0.0
    INTEGRAL Likelihood = -8.44 Transmembrane 37 - 53 ( 31 - 57)
    INTEGRAL Likelihood = -8.12 Transmembrane 75 - 91 ( 69 - 93)
    PERIPHERAL Likelihood = 2.76 200

```

-445-

modified ALOM score: 2.19
 icml HYPID: 7 CFP: 0.438

*** Reasoning Step: 3

5

----- Final Results -----

bacterial membrane --- Certainty=0.4376 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

10

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB12118 GB:Z99105 ycgQ [Bacillus subtilis]
 Identities = 100/290 (34%), Positives = 159/290 (54%), Gaps = 25/290 (8%)

15

Query: 9 MIRFLILAGYFELSMYLKLSGKLNQYINTHYTYLAYISMVLSFILAIVQLIIWVKNMKMH 68
 M R L+L G+ +L SG L +YIN Y YL++I++ L IL VQ +++K+ +
 Sbjct: 1 MFRLLVLMGFTFFFYHLHASGNLTKYINMKYAYLSFIAIFLLAILTAVQAYLFIKSPEKS 60

20

Query: 69 SHLHGKIA-----KSTSP-----MILVFPVLVGLLVPTVSILDSTTVSAKGYN 110
 H H + P ++ +FP++ G+ P +LDS+ V KG++
 Sbjct: 61 GHHDHDHDCGC HDHEHDHEQNKPFYQRYLIYVVFLFPLVSGIFFPIATLDSSIVKTGFS 120

25

Query: 111 FPLAAGSTGTVSQDGTRVQYLKPDTSYFTSSAYEKEMQKELKKYKGSGTLTTENYME 170
 F A S SQ QYL+PD S Y+ +Y+K+M++ KY +++T ++++++
 Sbjct: 121 FK-AMESGDHYSQ---TQYLRPDASLYYAQDSYDKQMKQLFNKYSSKKEISLTDFFLK 175

30

Query: 171 VMELIYLYPEQFMDRQIQYTGFVY-NEPKHEGYQFIFRGIIHCIA DSGVYGLLT-GNQ 228
 ME IY YP +F+ R I++ GF Y ++ F+ RFGIIHCIA DSGVYGL
 Sbjct: 176 GMETIYNPGEGFLGRTIEFHGFAYKGNAINKNQLFVLRFGIIHCIA DSGVYGLVEFPKD 235

35

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1145> which encodes the amino acid sequence <SEQ ID 1146>. Analysis of this protein sequence reveals the following:

Possible site: 60

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.33	Transmembrane	83 - 99 (74 - 101)
INTEGRAL	Likelihood = -6.21	Transmembrane	42 - 58 (39 - 62)

----- Final Results -----

bacterial membrane --- Certainty=0.4333 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

45

A related sequence was also identified in GAS <SEQ ID 9115> which encodes the amino acid sequence <SEQ ID 9116>. Analysis of this protein sequence reveals the following:

Possible cleavage site: 54

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL	Likelihood = -8.33	Transmembrane	75 - 91 (66 - 93)
INTEGRAL	Likelihood = -6.21	Transmembrane	34 - 50 (31 - 54)
PERIPHERAL	Likelihood = 2.76		

55

----- Final Results -----

bacterial membrane --- Certainty= 0.433 (Affirmative) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

60

An alignment of the GAS and GBS proteins is shown below:

Identities = 208/279 (74%), Positives = 244/279 (86%), Gaps = 1/279 (0%)

-446-

```

Query: 1 MFICGGNIMIRFLILAGYFELSMYLKLSGKLNQYINTHTYTLAYISMVLSFILAIQLII 60
      +F CGG +MIRFLILAGYFEL+MYL+LSGKL+QYIN Y+YLAYISM+LSFILA+VQL
Sbjct: 1 LFTCGGALMIRFLILAGYFELTMYLQLSGKLDQYINVRYSYLAYISMILSFILALVQLYT 60

5   Query: 61 WVKNMKMHSHLHGKIAKSTSPMILVFVPLVGLLVPTVSLDSTTVSAKGYNFPLAAGSTGT 120
      W+KN+K+HSHL GKIA+ TSP ILVFPVL+GLLVPTV+LDSTTVSAKGY FPLAAG++ T
Sbjct: 61 WMKNIKVHSHTLGKIARLTSPFILVFPVLIGLLVPTVTLDSSTTVSAKGYTFPLAAGASKT 120

10  Query: 121 -VSQDGTRVQYLPKPDTSYFTSSAYEKEMQKELKKYKGSGTLTTENYMEVMELIYLYP 179
      VS DGT +QYLKPDT S YFT SAY+KEM+EL KYKG +TITTE NYMEVMELIYLYP
Sbjct: 121 GVSDDGTTIQYLPKPDTSLYFTKSAYQKEMRQELHKYKGKKPVTTENYMEVMELIYLYP 180

15  Query: 180 EQFMDRQIQTGFVYNEPKHEGYQFIFRGFIIHCIADSGVYGLLTGNQKSYPDNTWVTV 239
      ++F+DR IQYTGFVYNEP H+ YQF+FRFGIIHCIA DSGVYGLLTGNQ SYP+NTW+TV
Sbjct: 181 DEFLLRDIDQYTGTVYNEPGHDNYQFLFRFGIIHCIA DSGVYGLLTGNQTSYPNNTWLTV 240

20  Query: 240 RGTIKSEYNQLLQQNLQVSKLHTEESRQVSKANNPYVYRVF 278
      +G + EY++ L+Q+LPVL + E Q + NNPyVYRVF
Sbjct: 241 KGRLHMEYDKNLEQHLPVLQLAEVHQTKEPNNPYVYRVF 279

```

SEQ ID 8560 (GBS235d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 14 & 15; MW 48.5kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 146 (lane 17 & 18; MW 23.4kDa), in Figure 150 (lane 15; MW 23kDa) and in Figure 182 (lane 5; MW 23kDa).

25 GBS235d-His was purified as shown in Figure 235, lane 6-7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 354

A DNA sequence (GBSx0385) was identified in *S.agalactiae* <SEQ ID 1147> which encodes the amino acid sequence <SEQ ID 1148>. This protein is predicted to be signal recognition particle (ftsY). Analysis of this protein sequence reveals the following:

```

Possible site: 57
>>> Seems to have no N-terminal signal sequence

35 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.3301(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

40 The protein has homology with the following sequences in the GENPEPT database:

```

>GP:BAB06205 GB:AP001515 signal recognition particle (docking
protein) [Bacillus halodurans]
Identities = 175/304 (57%), Positives = 227/304 (74%)

```

45 Query: 233 EKYNRSLLKTRGFSARLNALSNFRRVDEEFFEELEEMILISDVGVNATQLTEDLRYE 292
 EK+ L+KTR F+ ++N + +R VDE+FFEELEE+LI +DVGV L E+L+ E
Sbjct: 20 EKFKAGLEKTRDSFAGKMNDLVYKYRSVDEDFFEELEEILIGADVGVTVMDSLVEELKDE 79

50 Query: 293 AKLENAKKSEDLKRVIVEKLVEIYEKDGIYNEAINFQEGLTVMLFVGVNGVGKTTSIGKL 352
 + +N K S+D++ +I EKL E+ EK+G E GL+V+L VGVNGVGKTTSIGKL
Sbjct: 80 VRRQNIKDSKDIQPIISEKLAELLEKEGGETEVNLQPAGLSVILVVGVNGVGKTTSIGKL 139

Query: 353 AHQYKSQGKKVMLVAADTFRAGAVAQLVEWGRVDVPVVTGEEKADPASVVFDMEKAVA 412
 AH YK QGKKV+L A DTFRAGA+ QL WG R V V+ E +DPA+V+FD ++ A +
Sbjct: 140 AHMYKQQGKKVILAAGDTFRAGATEQLEVWGERAGVDVIKQSEGSDPAAVMFDAIQAAKS 199

Query: 413 QGVDVLLIDTAGRLQNKENLMAELEKIGRIIKRVPDAPHETLLALDASTGQNALSQAKE 472

-447-

+ D+L+ DTAGRLQNK NLM ELEK+ R+I R +P APHE L+ALDA+TGQNA+SQAK
 Sbjct: 200 READILICDTAGRLQNKVNLMKELKVKRVISREIPGAPHEVLIALDATTGQNAMSQAKT 259

5 Query: 473 FSKITPLTGLILTKIDGTAKGGVLAIRQELDIPVKFIGFGEKIDDIGEFNSEDFMRGLL 532
 F + T +TG+ILTK+DGTAKEGG+VLAIR ELDIPVKF+G GEKIDD+ F+SE F+ GL
 Sbjct: 260 FKETTDVTGIIILTKLDGTAKGGIVLAIRHELDIPVKFVGLGEKIDDQPFDSEQFVYGLF 319

Query: 533 EGIL 536

+ ++

10 Sbjct: 320 KDMV 323

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1149> which encodes the amino acid sequence <SEQ ID 1150>. Analysis of this protein sequence reveals the following:

Possible site: 60

15 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.4384 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

20 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 339/549 (61%), Positives = 404/549 (72%), Gaps = 46/549 (8%)

25 Query: 1 MGLFDRLFGHKKKDEKEPEIEAASESVVLEDEDSDVIDKEEGSNFSKESTLNRTSEVPVAEDD 60
 MGLFDRLFG K+ K E + E+++ E KEE S + E ++ + +
 Sbjct: 1 MGLFDRLFGKKETPKVAEEKLEENLLTE---TTQKEELSEKANEQ----DKIEAVQQE 51

30 Query: 61 SFLELERDTALSESHQPVTSIEHPLESEDTDEIPVKEDDSFLELEDRAKTKVADTSEVEN 120
 ++ + A S + P + ++ L E+T D + DT+E
 Sbjct: 52 ---DVSSEGAGSVENGPEAAVSNALVEETG-----DNSNHPSEDTNEF-- 92

35 Query: 121 VVPDSTTLSDNVSAKSEASFSDKEQLSDSQASDQFSETPLQEEMS--SGKTEVQTESEDT 178
 D T L VS S+++ S+ + L D +QF Q + S S E S++
 Sbjct: 93 -AADKTDLK--VSELQSQTASEPKDLVDQPVVEQFPTKQAAQADASNDSANEEAVDTSKEQ 149

40 Query: 179 SAADAFLADYYAKRKAIKEISSLNLST-----DESEFSEAQEVLQSQA--DTIK 227
 S++ + DYY ++ A+EK + + +T E++ S + E SQ++A DTI
 Sbjct: 150 SSSQQVMEDYYRRKAALEKSLQEKAATAVPVMPEEVQPQENQASTSAEA-SQNKATHDTIP 208

Query: 228 AESQEEKYNRSLKKTRTGF SARLNAFLSNFRVDEEFFEELEEMILSDVGVN VATQLTE 287
 E+ +EKY RSLKKTRTGF SARLN+F +NFRVDEEFFE+LEEMILSDVGV+VAT LTE
 Sbjct: 209 -ETDQEKYKRSLLKTRTGF SARLNSFFANFRVDEEFFEDEMLILSDVGVHVATTLTE 267

45 Query: 288 DLRYEAKLENACKSEDILKRVIVEKLVEIYEKDGIYNEAINFQEGLTVMLFVGVNGVGKTT 347
 +LRYEAKLENACK + LKRVIVEKLV+IYEKDGIYNEAIN+Q+GLTVMLFVGVNGVGKTT
 Sbjct: 268 ELRYEAKLENACKPDALKRVIVEKLVDIYEKDGRYNEAINYQDGLTVMLFVGVNGVGKTT 327

50 Query: 348 SIGKLAHQYSQGKKVMLVAADTFRAGAVAQLVEWGRDVDPVVTGEEKADPASVVF DGM 407
 SIGKLA++YK +GKKVMLVAADTFRAGAVAQLVEWGRDVDPV+TG EKADPASVVF DGM
 Sbjct: 328 SIGKLAYRYKQEGKKVMLVAADTFRAGAVAQLVEWGRDVDPVITGPEKADPASVVF DGM 387

Query: 408 EKAVAQGV D VLLIDTAGRLQNKENLMAELEKIGRIIKR VVPDAPHETLIALDASTGQNAL 467
 EKAVA+GVD+LLIDTAGRLQNKENLMAELEK+GRIIKR+PDAPHETLIALDASTGQNAL
 Sbjct: 388 EKAVAKGV D ILLIDTAGRLQNKENLMAELEKMGRIIKR VLPDAPHETLIALDASTGQNAL 447

55 Query: 468 SQAKEFSKITPLTGLILTKIDGTAKGGVLAIRQELDIPVKFIGFGEKIDDIGEFNSEDF 527
 SQAKEFSKITPLTGLILTKIDGTAKGGVLAIRQELDIPVKFIGFGEK+DDIGEF+SEDF
 Sbjct: 448 SQAKEFSKITPLTGLILTKIDGTAKGGVLAIRQELDIPVKFIGFGEKVFDDIGEFHSEDF 507

60 Query: 528 MRGLLEGIL 536
 M+GLLEGIL
 Sbjct: 508 MKGLLEGIL 516

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 355

A DNA sequence (GBSx0386) was identified in *S.agalactiae* <SEQ ID 1151> which encodes the amino acid sequence <SEQ ID 1152>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3592 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAA62048 GB:L10328 f270 [Escherichia coli]
 Identities = 101/273 (36%), Positives = 160/273 (57%), Gaps = 10/273 (3%)

Query: 4 IKILALLDGTLFTTDKKVSEENKVALKAAREKGIGKVITGRPLKAIGNLLEDLELVSD 63
 IK++A+D+DGTL D +S K A+ AAR +G+ VV+TTGRP + N L++L +
 Sbjct: 3 IKLIAIDMDGTLLLPDHTISPRAVKNAIAAARARGVNVLTTGRPYAGVHNLYLKEHMEQP 62

Query: 64 EDYSITFNGGLVQQNT-GKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTVYS---I 118
 DY IT+NG LVQ+ G +A+TA++ + + + +VG L T+Y+ I
 Sbjct: 63 GDYCITYNGALVQKAADGSTVAQTALSYDDYRXLEKLSREVGSHFHALDRRTLYTANRDI 122

Query: 119 ANKGHHHSQYHLANPLLEFTIEVDDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRLKVDY 178
 + H + PL+ F E E++ + + K++ ++ LDQ IA++P +K Y
 Sbjct: 123 SYYTvhESFVATIPLV-FCEA--EKMDPNTQFLKVMIMDEPAILDQAIARIPQXVKEKY 178

Query: 179 EMPKSRDIILELMPKGVHKAVGLELLITKHLGLDSSQVMAMGDEANDLSMILEWAGLGVAMA 238
 + KS LE++ K V+K G++ L LG+ ++MA+GD+ ND++M+E+AG+GVAM
 Sbjct: 179 TVLKSAPYFLEILDKRVNKGTGVKSLADVLGIKPEEIMAIGDQENDIAMIEYAGVGVAMD 238

Query: 239 NGIPEAKAKATTICNNNDESGVAAEAIKYIILS 271
 N IP K +A T +N E GVA AI KY+L+
 Sbjct: 239 NAIPSVKEVANFVT-KSNLEDGVAFIAIEKYVLN 270

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1153> which encodes the amino acid sequence <SEQ ID 1154>. Analysis of this protein sequence reveals the following:

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3502 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 180/273 (65%), Positives = 218/273 (78%), Gaps = 1/273 (0%)

Query: 3 DIKILALLDGTLFTTDKKVSEENKVALKAAREKGIGKVITGRPLKAIGNLLEDLELVSD 62
 +I+ILALLDGTL+ T+K V++ NK AL AAREKG+KVITGRPLKAIGNLLE+L+L+
 Sbjct: 2 NIRILALLDGTLNYTEKIVTDANKKALAAAREKGVKVVITGRPLKAIGNLLEELLDLL 61

Query: 63 DEDYSITFNGGLVQQNTGKILAKTAMTRQEVEDIHEELYQVGLPTDILSEGTVYSIANK- 121
 +DYSITFNGGLVQ+NTG++L K++++ +V I + L VGLPTDI+S G VYSI +K
 Sbjct: 62 HDDYSITFNGGLVQRNTGEVLDKSSLSDQVCQIQQQALEAVGLPTDIISGGDVYSIPSKD 121

Query: 122 GHHSQYHLANPLLEFTIEVDDLEQVPKDVVYNKIVSVIDATYLDQQIAKLPDRLKVDYEMF 181

-449-

G HSQYHLANPLL FIEV + ++PKD+ YNKIV+V D +LDQQI KL L D+E F
 Sbjct: 122 GRHSQYHLANPLLTIEVTVAELPKDITYNKIVTVTDPFLDQQIIKLSPLSFEDFEAF 181

5 Query: 182 KSRDIILELMPKGVHKAVGLELLTKHLGLDSSQVMAMGDEANDLSMWEAGLGVAMANGI 241
 KSRDII E+MPKG+ KA GL LL +HLGLD+ VMAMGDEAND +MLEWAGLGVAMANG+
 Sbjct: 182 KSRDIIFEIMPKGIDKAFCGLNLLCQHGLDARHVMAMGDEANDFAMLEAGLGVAMANGV 241

10 Query: 242 PEAKAIKATTICNNDESGVAEAIGKYILSEEN 274
 AKA A A T NDESGVAEEA+ +IL EE+
 Sbjct: 242 SGAKADADAVTTLTNDESGVAEAVKTFILEEES 274

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 356

15 A DNA sequence (GBSx0387) was identified in *S.agalactiae* <SEQ ID 1155> which encodes the amino acid sequence <SEQ ID 1156>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

20 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4648 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

25 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA35556 GB:D90723 Hypothetical 30.2 kd protein in idh-deoR
 intergenic region. [Escherichia coli]
 Identities = 91/264 (34%), Positives = 146/264 (54%), Gaps = 4/264 (1%)

30 Query: 2 IKLVATMDGTFLDENGYDKKRLANVLKKFKEQGIVFTAASGRSLLSLEQLFADFRDQM 61
 IKL+A DMGTFL + TY+++R ++ K QGI F ASG L F + +++
 Sbjct: 4 IKLIAVDMGTFLSDQKTYNRERFMAQYQQMKAQGIRFVVASGNQYYQLISFFPEIANEI 63

35 Query: 62 AFIAENGSAAVLFNRILAYEQHLSREQYLDIIDHLSKSPYMEENNEYVLSGKDAGAYILSDAN 121
 AF+AENG V + + LS++ + +++HL P + E + GK+ AY L +
 Sbjct: 64 AFVAENGWWVSEGKDVNGELSKDAFATVVEHLLTRPEV--EIIACGKNSAYTLKKYD 120

Query: 122 PDYIEFITHYYDNLQKVSHFEDVDDIIFKVIANFTTEETVRQAEEWVNQAI-PYATAVTTG 180
 YY L+ V +F++++DI FK N ++E + Q ++ +++AI +V TG
 40 Sbjct: 121 DAMKTVAEMYYHRLEYVDNFNDNLEDIFFKFGLNLSDELIPQVQKALHEAIGDIMVSHTG 180

Query: 181 FKSIDIIILSSVNVKRNGLEHLCEQYGIRAEEVLSFGDNINDLEMLEWSGKAIATEPARPEV 240
 SID+I+ V+K NGL L + +GI EV+ FGD ND+EML +G + A ENA V
 Sbjct: 181 NGSIDIIIPGVHKANGLRQLQKLWGIDDSEVVFGDGGNDIEMPLRQAGFSFAMENAGSAV 240

45 Query: 241 KEIADCIIGHHNNQAVMAYLESV 264
 A G +N + V+ ++ ++
 Sbjct: 241 VAAAQYRAGSNREGVLDVIDKVL 264

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1157> which encodes the amino acid sequence <SEQ ID 1158>. Analysis of this protein sequence reveals the following:

Possible site: 49
 >>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3401 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

60 An alignment of the GAS and GBS proteins is shown below:

-450-

Identities = 138/265 (52%), Positives = 193/265 (72%), Gaps = 1/265 (0%)

```

Query: 1 MIKLVATDMGTFLDENGTYDKKRLANVLKKFKEQGIVFTAASGRSLLSLEQLFADFRDQ 60
      MIKL+ATDMGTFPL E+GTY++++LA +L K E+GI+F +SGRSLL+++QLF F DQ
5   Sbjct: 1 MIKLIATDMGTFLAEDGTYNQEQLAALLPKLAEKGILFAVSSGRSLLAIDQLFEPFLDQ 60

Query: 61 MAFIAENGSAAVLFNRLAYEQHLSREQYLDIIDHLSKSPYMEMNEYVLSGKDGYILSDA 120
      +A IAENGSS + + +++EQY ++ + +P+ V SG+ AYIL A
10  Sbjct: 61 IAVIAENGSVVQYRGEILFADMNTKEQYTEAKKILANPHYVETGMVFSGQKAAYILKGA 120

Query: 121 NPDYIEFITHYYDNLQKVSHFEDVD-DIIFKVTAANFTEETVRQAEEVNQAI PYATAVTT 179
      + +YI+ HYY N++ ++ FED++ D IFKV+ NFT TV + +W+NQA+PYATAVTT
      Sbjct: 121 SEEYIQQTKHYYANVKVINGFEDMENDAIFKVSTNFTGHTVLEGSDWLNLQALPYATAVTT 180

15  Query: 180 GFKSIDIILSSVNKRNGLEHLC EQYQGIRAEVLSFGDNINDLEMLEWSGKAIATEENARPE 239
      GF SIDIIL VNK G+EHLCL+ GI+ E ++FGDN ND +MLE++G+AIATEENARPE
      Sbjct: 181 GFDSIDIILKEVNKGFGMEHLCQALGIKKAETIAFGDNFNDYQMELFAGRAIATEENARPE 240

20  Query: 240 VKEIADCIIIGHHNQNQAVMAYLES 264
      +K I+D +IGH N+ AV+ YL+ +V
      Sbjct: 241 IKVISDQVIGHCMDGAVLTYLKGLV 265

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 357

A DNA sequence (GBSx0388) was identified in *S.agalactiae* <SEQ ID 1159> which encodes the amino acid sequence <SEQ ID 1160>. Analysis of this protein sequence reveals the following:

```

Possible site: 18
>>> Seems to have no N-terminal signal sequence
30
----- Final Results -----
      bacterial cytoplasm --- Certainty=0.2428 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
35

```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

40 Example 358

A DNA sequence (GBSx0389) was identified in *S.agalactiae* <SEQ ID 1161> which encodes the amino acid sequence <SEQ ID 1162>. This protein is predicted to be p115 protein (smc). Analysis of this protein sequence reveals the following:

```

Possible site: 55
45  >>> Seems to have no N-terminal signal sequence
      INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)

----- Final Results -----
      bacterial membrane --- Certainty=0.2996 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
50

```

A related GBS nucleic acid sequence <SEQ ID 9713> which encodes amino acid sequence <SEQ ID 9714> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolog
 [Bacillus subtilis]
 Identities = 458/1193 (38%), Positives = 728/1193 (60%), Gaps = 27/1193 (2%)

5 Query: 1 MFLKEIEMQGFKSFADKTKEFDQGVTAVVGPNGSGKSNITESLRWALGESEAKSLRGKK 60
 MFLK +++ GFKSFA++ V+F +GVTAVVGPNGSGKSNIT+++RW LGE SA+SLRGKK
 Sbjct: 1 MFLKRLLDVIGFKSFAERISVDFVKGVTAVVGPNGSGKSNITDAIRWLGEQSARSLRGKK 60

10 Query: 61 MPDVIFACTENRKPLNQAQVSVTLDNSDHFIFIENIADEVRVERRIFRNGDSEYLIDGRKVR 120
 M D+IFAG+++RK LN A+V++TLDN DHF+ EV V RR++R+G+SE+LI+ + R
 Sbjct: 61 MEDIIIFAGSDSRKLNLAEVTLTLDDHFLPIDFHEVSVTRRVYRSGESEFLINNQPCR 120

15 Query: 121 LRDIHDLFMDTGLGRDSFSIISQGRVEAIFNSKPEERRAI FEEAAAGVLKYKTRKKE TQSK 180
 L+DI DLFMD+GLG+++FSIISQG+VE I +SK E+RR+I FEEAAAGVLKYKTRKK+ ++K
 Sbjct: 121 LKDIIDLFMDSGLGKEAFSIISQGKVEEILSSKAEDRRSIFEEAAAGVLKYKTRKKKAENK 180

20 Query: 181 LETOGNLDRLEDIYELDMQVOPLEKOASI AKRFLVLDEERQGLHLSILIEDILOHSD 240
 L +TQ NL+R+EDI++EL+ QV+PL+ QASI AK +L +E + + + + DI +
 Sbjct: 181 LFETQDNILNRVEDILHELEGQVEPLKIQASI AKDYLEKKKELEHVEIALTAYDIEKLHGK 240

25 Query: 241 LTTEVEKLLTVRKELATYYQQRQSLDENQSLKQKRHHLSEEIEAKQLALLDVTKLKSDL 300
 +T++EK+ ++E + E + + + K L E + Q LL ++ L
 Sbjct: 241 WSTLKEKVQMAKEEEELAES AISA EAKIE DTRDKI QALDES VNELQVLLVTSEELEKL 300

30 Query: 301 ERQIDLIRLESNQKAEKKEEAGQRLAEL EIKAKDCSDQITQKNIELTTLSEKIAQIRSEI 360
 E + + + + + + E+ + + + + K ++++++ TL ++ Q+R+++
 Sbjct: 301 EGRKEVLKERKKNAVNQNEQLEEAIVQFQQKETV LKEELSKQEA FVETLQAEVKQLRAQV 360

35 Query: 361 VSTESSLERFSTNPDQIEKLREDFTLQMEEADTSNALTALLADIE NQKQASQAKSQEI 420
 + +L + N ++ IE+L+ D+ L+ +A N L LL D +Q + + +
 Sbjct: 361 KEKQQALSLHNENVEEKIEQLKSDYFELLNSQASIRNEL-QLLDDQMSQS AVTLQR LADN 419

40 Query: 421 QEVS KNLEV LKSNAKVALE-RFEAAKKNV RQLLSHYQDLGQT LQNLEGEYKNQ QSI LF DH 479
 E S K A E F ++ + + Y+D+ + + + + Y+ + S L+
 Sbjct: 420 NEKHLQERHDISARKAACETEFARIEQEIHSQVGAYRDMQTKYEQKKRQYEKNESALYQA 479

45 Query: 480 LDEIKSKQARISSLESILKNHNSNFYAGVKS V LQAKDQLGGIIGAVSEHLSFDKHYQT ALE 539
 ++ + + + LE++ + S FY GVK VL+AK++LGGI GAV E +S ++ Y+TA+E
 Sbjct: 480 YQYVQQQARSKKDMLETMQGDFSGFYQGVKEVLAKERLGGIRGA VLELISTEQKYETAIE 539

50 Query: 540 I ALGGSSQHII VEDESAAKRSIAFLKKNRQGRATFLPLTTIKPRELAQH YLSKLQSSQGF 599
 I ALG S+QH++ +DE +A+ + +I +LK+N GRATFLPL+ I+ R+L F
 Sbjct: 540 I ALGASAQHVVT DDEQSARKAIQYLKQNSFGRATFLPLSVIRDRQLQSRDAETAARHSSF 599

55 Query: 600 LGIASELV TYDQLSNI FKNNLGLTAIFDTVDNANVAARQ LNYQVR LVTLDGT ELRPGGS 659
 LG+ASELVT+D ++ +N LG I + + AN A+ L ++ R+VTL+G + PGGS
 Sbjct: 600 LGVASELVT FDPAYRSVIQNL LGTV LITEDLKGANELAKL LGHRYRIVTLEG DVVNP GGS 659

60 Query: 660 YSGGANRQNN TVFI--KPELDNLKKELKQ AQS KQ L I Q E KEV AT LLEQLKEKQ ET LAQ LKN 717
 +GGA ++ N + EL++ + K L + + K + E+EV TL + + + + LA L+
 Sbjct: 660 MTGGAVKKKKNSLLGRSRELE DVT KRLAEME EKT ALLE QEV KTLKHSI QDMEK KLA DLRE 719

65 Query: 718 DGEQARLEEQRADIEYQQLSEK LADLN KLYNGLQ LSSG ALEQTT SENE --K NRLEKELEQ 775
 GE RL++Q + +L ++N AL ++ E + K + LE+EL
 Sbjct: 720 TGEGLRLKQ QDV KGQ L YELQVAE KNINT HLELYDQ EK SALS EDEER KVR KR KLE EELSA 779

70 Query: 776 FAIKKEELT TSIAQ I KEDK DS I QEK VN NL TLLSEA QL EER DLLNEQ KFER ANCTRL--- 832
 + K ++L I ++ + K + +L+ L+E ++ K E N RL
 Sbjct: 780 VSEKMKQLEEDIDRLT KQK QTOS STKESLSN ELTEL KIAA AKKEQACK GEEDNL ARLKKE 839

75 Query: 833 ----EITLSEI KRD ISN LQL TLLSHQD SQLKEELPRI EK QL LQVN N RREN DEE KL VSLRF 888
 E+ L E K D+S L + +S S E++L + + ND+ K + L
 Sbjct: 840 LTETELA LKEAKEDLSFLT SEMSS STSG-----E EK LEEAAKHKLNDKT KTI ELIA 890

80 Query: 889 ELEDCEA ALDDIA ASLAKE GQKNESLIR QQAQL---ESQCEQLS QQLMIFS RQL SEDYQ 944
 D L + +E ++ + L +Q+ L E + ++ +L + L E+Y

-452-

Sbjct: 891 LRRDQRRIKLOHGLDTYERELKEMKRKYQKTTLLKDEEVKLGRMEVELDNLLQYLREEYS 950

Query: 945 MTLDEAKVKANVLEDILMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLTFNLNTQRDD 1004
++ + AK K + D AR+++K ++ I+ LG VN+ +I +FE V+ER FL+ Q++D

5 Sbjct: 951 LSFEGAKEKYQLETDPEEARKRVLKLIKATEELGTVNLGSIDEFERNERYKFELSEQKED 1010

Query: 1005 LVHAKNLLLETITDMDEEVTRFKSTFEAIRHSFKETFVQMFGGGSADLILTE-GDLLSA 1063
L AKN L + I +MD+E+ RF TF IR F + F +FGGG A+L LT+ DLL +

10 Sbjct: 1011 LTEAKNTLFQVIEEMDEEMTKRFNDTFVQIRSHFDQVFRLSGGGRAELRLTDPNLLHS 1070

Query: 1064 GVDISVQPPGKKIQSLNLSMCGEKALSALLFAIIRVKTIPFVILDVEEAALDEANVKR 1123
GV+I QPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEEAALDEANV R

Sbjct: 1071 GVEIIAQPPGKKLQNINLLSGGERALTAIALLFSILKVRPVFCVLDVEEAALDEANVFR 1130

15 Query: 1124 FGDYLNRFDKSSQFIVVTHRKGTMQAADSIFYGVTMQESGVSKIVSVKLKEAQE 1176
F YL ++ +QFIV+THRKGTM AD +YGVTMQESGVSK++SVKL+E +E

Sbjct: 1131 FAQYLKKYSSDTQFIVITHRKGTMEEADVLYGVTMQESGVSKVISVKEETKE 1183

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1163> which encodes the amino acid
20 sequence <SEQ ID 1164>. Analysis of this protein sequence reveals the following:

Possible site: 15

>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -4.99 Transmembrane 1092 -1108 (1088 -1110)

25 ----- Final Results -----

bacterial membrane --- Certainty=0.2996 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the databases:

>GP:CAB13467 GB:Z99112 chromosome segregation SMC protein homolog
[Bacillus subtilis]
Identities = 441/1192 (36%), Positives = 729/1192 (60%), Gaps = 25/1192 (2%)

35 Query: 1 MFLKEIELEGFKSFADKTKIEFDKGVTAVVGPNNGSGKSNTESLRWLGESSIONLRRGGK 60
MFLK +++ GFKSFA++ ++F KGVTAVVGPNNGSGKSNT+++RW LGE SA++LRRGGK
Sbjct: 1 MFLKRLDVIGFKSFAERISVDFVKGVTAVVGPNGSGKSNTDAIRWLGEQSARSLRRGGK 60

40 Query: 61 MPDVIFAGTQNRRNPLNYAKVAVVLDNSDHFIKTAKKERVERHIYRNQGDSDYLIDGRKVR 120
M D+IFAG+ +R LN A+V + LDN DHF+ E+ V R +YR+G+S++LI+ + R
Sbjct: 61 MEDIIIFAGSDSRKRLNLAEVTLTLDNDDHFPLIDFHEVSVRTRVYRSGESEFLINNQPCR 120

45 Query: 121 LRDIDHDFMDTGLGRDSFSIISQGRVEEIFNSKPEERRAIFEEAAGVLKYKTRKKETQIK 180
L+DI DLFMD+GLG++FSIISQG+VEEI +SK E+RR+IFEEAAGVLKYKTRKK+ + K
Sbjct: 121 LKDIIDLFMDSGLGKEAFSIIISQGKVEEILSSKAEDRRSIFEAAAGVLKYKTRKKKAENK 180

50 Query: 181 LNQTQDNLDRILEDIIYELDTQLAPLEKQAKVAKQFELDANRKQLQLDILVKDIDIAQER 240
L +TQDNL+R+EDI++EL+ Q+ PL+ QA +AK +LE + +++ + DI+ +
Sbjct: 181 LFETQDNLNRVEDILHELEGQVPLKIQASIAKDYLEKKLEHVEIALTAYDIEKLHGK 240

55 Query: 241 QTQDTEALALQQDLASYYAKRQSMEEDYQKFKQKKQVLSQESDQTQTTLELTKLIADL 300
+ E + +++ + + E + + K Q L + ++ Q LL ++ + L
Sbjct: 241 WSTLKEKVQMAKEEELAESAIASAKEAKIEDTRDKIQALDESVNELQQVLLVTSEELEKL 300

60 Query: 301 EKQIELVKLESQAEKKAAEAKKHLQLQEQLDGFQAEKQCTEQLLH-----IDQQL 353
E + E++K E+K A ++ EQL+E + FQ +E E+L + ++
Sbjct: 301 EGRKEVLK-----ERKKNAVQNQEQLERAIVQFQQKETVLKEELSKQEAVFETLQAEV 353

65 Query: 354 CDVKQQLNELSNALERFSSDPDQLMETLREEFVLLMQKEAALSNQLTALKAHLDKEQAR 413
++ Q+ E AL + + ++ +E L+ ++ L+ +A++ N+L L + +
Sbjct: 354 KQLRAQVKKEKQQLSLHNENVEEKIEQLKSDYFELLNSQASIRNELQLLDDQMSQSAVTI 413

Query: 414 QHKAQEYQLLVTKLDQLNDESQKAQAHYKAQKEQVEMLLQNYQEGDKRVQELERDYQLNQ 473
Q A + + + ++ + + + +++++ + Y++ + ++ +R Y+ N+

Sbjct: 414 QRLADNNEKHLQERHDISARKAACETEFARIEQEETHSQVGAYRDMQTKYEQKKRQYEKNE 473

Query: 474 ERLFDLLDQKKGKEARKASLESIQKSHSQFYAGVRAVLQSQKKLGGIIGAVSEHLSFDSD 533
L+ + ++K LE++Q S FY GV+ VL+++++LGGI GAV E +S +

5 Sbjct: 474 SALYQAYQYVQQARSKKDMLETMQGDFSGFYQGVKEVLAKERLGGIRGAVLELISTEQK 533

Query: 534 YQTAIEVALGANSQHIIVTDEAAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHQL 593
Y+TA+E+ALGA++QH++ DE +A++AI YLK+N GRATFLPL+ I+ R L

10 Sbjct: 534 YETAIEIALGASAQHVVTDDQSARKAIQYLQNSFGRATFLPLSVIRDRLQLSRDAETA 593

Query: 594 ATCEGYLGTAEISLIRYDDSLSAIIQNLSSSTAIFETIDQANIAARLLGYKVRIVTLDGTE 653
A +LG A L+ +D + ++IQNLL + I E + AN A+LLG++ RIVTL+G

Sbjct: 594 ARHSSFLGVASELVTFDPAYRSVIQNLGTVLITEDLKGANELAKLLGHRYRIVTLEGDV 653

15 Query: 654 LRPCCGSFSGGANRQSNNTTFI--KPELEQISEELTRIVEQLKITEKEVAALQSDLIAKKEE 711
+ PGGS +GGA ++ N + + ELE +++ L + E+ + E+EV L+ + + +

Sbjct: 654 VNPGGSMTGGAVKKNNNSLLGRSRELEDVTKRLAEMEKTALLEQEVKTLKHSIQDMEKK 713

20 Query: 712 LTOLKLAGDQARLAEQ--RAQMAYQOLQEKQEDSKALLAALDQSQTTHSDESLLAEQARI 769
L L+ G+ RL +Q + Q+ EK ++ L ++S + SDE + ++

Sbjct: 714 LADLRETGEGLRLKQQDVKGQLYELQVAEKNINTHLELYDQEKSALSESDEERKVRKRKL 773

Query: 770 EEALTAIAKKKNALTCDIDDIKENKDLIROQTKONIHQALSQARLQERDILLNEKKFEQANQ 829

EE L+A+++K L DID + + K +++ L++ ++ K E+ N

25 Sbjct: 774 EEELSAVSEKMKQLEEDIDRLTKQKQTQSSTKESLSNELTELKIAAAKKEQACKGEEDNL 833

Query: 830 SRLRTQLKQCQQNILKLESILNNNVSDSIQRLPQWQKQLQDATEHKSGAQKRLVQLRFE 889

+RL+ +L + + + + L+ S+ S +++L++A +HK + + ++L

30 Sbjct: 834 ARLKKELTETELALKEAKEDLSFLTSEMSSS--TSGEEKLEAAKHKLNDKTKTIELIAL 891

Query: 890 IEDYEARLEETAEKITKESEKNDTFIRRQTKL---ETHLEQVANRLRAYAKSLSEDQFM 945

D +L+ + +E ++ +++T L E I ++ L + L E++ +

Sbjct: 892 RRDQRRIKLOHGLDTYERELKEMKRLYKQKTTLLKDEEVKLRMVEVELDNLLQYLREEYSL 951

35 Query: 946 TLADAKEVTNSIDHLESAKEKLHHLQKTIRALGPINSAINQYEEVHERLTFLTSQKTDL 1005

+ AKE E A++++ ++ I LG +N +I++E V+ER FL+ QK DL

Sbjct: 952 SFEGAKEKYQLETDPEEAKRVRVKKLIKLAIEELGTVNLGSIDEFERNERYKFLSEQKEDL 1011

40 Query: 1006 TKAKNLLLETINSMDEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTE-TDLLSAG 1064

T+AKN L + I MD E+ RF TF I+ F + F +FGGG A+L LT+ DLL +G

Sbjct: 1012 TEAKNTLFQVIEEMDEEMTKRFNDTFVQIRSHFDQVFRSLFGGGRAELRLTDPPNDLLHSG 1071

Query: 1065 IEISVQPPGKKIQSLNLMMSGGEKALSALALIFAIIRVKTIPFVILDEVEAALDEANVKRF 1124

+EI OPPGKK+Q+LNL+SGGE+AL+A+ALLF+I++V+ +PF +LDEVEAALDEANV RF

45 Sbjct: 1072 VEIIAQPPGKKLQNLLNLSSGERALTAIALLFSILKVRPVPCVLDVLEVEAALDEANVFRF 1131

Query: 1125 GDFLNRFDKDSQFIVVTHRKGTMMAADSIFYGITMQESGVSKIVSVKLKEAQE 1176

+L ++ D+QFIV+THRKGTM AD +YG+TMQESGVSK++SVKL+E +E

Sbjct: 1132 AQYLKKYSSDTQFIVITHRKGMEEADVLYGVTMQESGVSKVISVKLEETKE 1183

50

An alignment of the GAS and GBS proteins is shown below:

Identities = 732/1179 (62%), Positives = 911/1179 (77%)

55 Query: 1 MFLKEIEMQGFKSFADKTKVEFDQGVTAVVGPNGSGKSNITESLRWALGESSAKSLRGKK 60
MFLKEIE++GFKSFADKTK+EFD+GVTAVVGPNGSGKSNITESLRWALGESSAK+LRGGK

Sbjct: 1 MFLKEIELEGFKSFADKTKIEFDKGVTAVVGPNGSGKSNITESLRWALGESSAKNLRGKK 60

Query: 61 MPDVIFAGTENRKPLNYAQSVTLDNSDHFIENIADEVRVERRIFRNGDSEYLIDGRKVR 120

MPDVIFAGT+NR PLNYA+V+V LDNSDHFI+ E+RVER I+RNGDS+YLIDGRKVR

60 Sbjct: 61 MPDVIFAGTQNRNPPLNYAKVAVVLNDNSDHFIKTAKEIRVERHIYRNGDSDYLIDGRKVR 120

Query: 121 LRDIHDLFMDTGIGRDSFSIISQGRVEAIFNSKPEERRAIFEAAAGVLKYKTRKKETQSK 180

LRDIHDLFMDTGIGRDSFSIISQGRVE IFNSKPEERRAIFEAAAGVLKYKTRKKETQ K

65 Sbjct: 121 LRDIHDLFMDTGIGRDSFSIISQGRVEEIFNSKPEERRAIFEAAAGVLKYKTRKKETQIK 180

Query: 181 LEQTQGNLDRLEDIIYELDMQVQPLEKQASIARKFLVLDEERQGLHLSILIEDILQHQSD 240

L QTQ NLDRLEDIYELD Q+ PLEKQA +AK+FL LD R+ L L IL++DI Q

-454-

5 Subjct: 181 LNQTQDNLDRLEDIYELDTQLAPLEKQAKVAKQFLELDANRKQLQDILVKDIDIAQER 240
 Query: 241 LTTVEEKLLTVRKELATYYQQRSLEDENQSLKQKRHHLSEEIEAKQLALLDVTKLSDL 300
 T E L +++LA+YY +RQS+E++ Q KQK+ LS+E + Q LL++TKL +DL
 Sbjct: 241 QTKEDEALALQDLASYYAKRSMEEDYQKFQKKQVLSQESDQTQTTLELTKLIADL 300

10 Query: 301 ERQIDLIRLESNQKAEEAQRLAELEIKAKDCSDQITQKNEITLSEKIAQIRSEI 360
 E+QI+L++LES Q+AEKK EA + L +L+ + + Q +L + ++ ++ ++
 Sbjct: 301 EKQIELVKLESQEAEEKAEEAKKHLEQLQEQLDFQAEKQCTEQLLHQDQLCDVQQL 360

15 Query: 361 VSTESSLERFSTNPQIIKEIUREDFVTLMQEEADTSNALTALEADENQKQASQAKSOEI 420
 ++LERFS++PDQ++E LRE+FV LMQ+EA SN LTAL A ++ +KQA Q K+QE
 Sbjct: 361 NELSNALERFSSDPDQLMETLREEFVLLMQKEAALSNLALKAHLDKEQARQHKAQEY 420

20 Query: 421 QEVSKNLEVLKSNAKVALERFEAAKKNVRQLLSHYQDLGQTLQNLEGEYKNQQSILFDHL 480
 Q + L+ L ++ A ++A K+ V LL +YQ+ + +Q LE +Y+ Q LFD L
 Sbjct: 421 QLLVTKLDQLNDESQKAQAHYKAQKEQVEMLLQNYQEGDKRVQELRDYQLNQERLF DLL 480

25 Query: 481 DEIKSKQARISSLESILKNHSNFYAGVKSVLQAKDOLGGIIGAVSEHLSFDKHYQTALEI 540
 D+ K K+AR +SLESI K+HS FYAGV++VHQ++ +LGGIIGAVSEHLSFD YQTALE+
 Sbjct: 481 DQKKKGKEARKASLESTIQKSHSQFYAGVRAVLQSQKKLGGIIGAVSEHLSFDSDYQTALEV 540

30 Query: 541 ALGGSSQHIIVEDESAARKSIAFLKKNRQGRATFLPLTTIKPRELAQHYLSKLQSSQGL 600
 ALG +SQHIIIV DE+AAKR+IA+LKKNRQGRATFLPLTTIK R L++HY +L + +G+L
 Sbjct: 541 ALGANSQHIIVTDEAAKRAIAYLKKNRQGRATFLPLTTIKARSLSEHYHRQLATCEGYL 600

35 Query: 601 GIASELVTDQRLLSNIFKNNLGLTAIFDTVDNANVAARQLNYQVRLVTLGDTELPGGSY 660
 G A L+ YD LS I +N L TAIF+T+D AN+AAR L Y+VR+VTLGDTELPGGS+
 Sbjct: 601 GTAESLIRYDDSLSAIIQNLSTAIFETIDQANIAARLLGYKVRIVTLGDTELPGGSF 660

40 Query: 661 SGGANRQNNTVFIKPELDNLKKELKQAKQSKQLIQEKEVATLLEQLKEKQETLAQLKNDGE 720
 SGGANRQ+NT FIKPEL+ + +EL + + I EKEVA L L K+E L QLK G+
 Sbjct: 661 SGGANRQSNITTFIKPELEQISEELTRLVQQLKITEKEVAALQSDLIAKKEELTQLLAGD 720

45 Query: 721 QARLEEQRADIEYQQLSEKLAIDLNLKLYNGLQLSSGALEQTTSENEKRNLEKELEQFAIKK 780
 QARL EQRA + YQQL EK D L L S + E+ R+E+ L A KK
 Sbjct: 721 QARLAEQRQAMYQQLQEKQEDSKALLAALDQSQTTHSDESLLAEQARIEEALTAIAKKK 780

50 Query: 781 EELTTSIAQIKEKDSDIQEKNVNLLTLLSEAQLEERDILLNEQKFERANCTRLEITLSEIK 840
 LT I IKE+KD I++K N+ LS+A+L+ERDLLNE+KFE+AN +RL L + +
 Sbjct: 781 NALTCDIDDIKENKDLIRQKTQNTIQALSQARLQERDILLNEKKFEQANQSLRTQLKQCQ 840

55 Query: 841 RDISNLQTLLSHQDSQLDKEELPRIEKQLLQVNNRRENDEEKLVSLRFELEDCEAALDDL 900
 ++I L+++L++ SQ + LP+ +KQL + + +LV LRFE+ED EA L++
 Sbjct: 841 QNLLKLESILNNNVSDSIQRLPOWQKQLODATEHKSGAQKRLVQLRFETIEDYEARLEET 900

60 Query: 901 AASLAKEGQKNESLIRQQAQLESQCEQLSQQLMIFSRLQSEDYQMTLDEAKVKANVLEDI 960
 A + KE +KN++ IR+Q +LE+ EQ++ +L + + LSED+QMTL +AK N ++ +
 Sbjct: 901 AEKITKESEKNDTFIRROTKLETHLEQVANRLRAYAKSLSEDQFQMTLADAKEVTNSIDHL 960

65 Query: 961 LMAREQLKSLQAKIKALGPVNIDAIAQFEEVHERLTFILNTQRDDLVHAKNLLLETITMD 1020
 A+E+L LQ I+ALGP+N DAI Q+EEVHERLTFL +Q+ DL AKNLLLETI MD
 Sbjct: 961 ESAKEKLHHLQKTIRALGPINSAINQYEEVHERLTFILTSQKTDLTKAKNLLLETINSM 1020

70 Query: 1021 DEVKTRFKSTFEAIRHSFKETFVQMFGGGSADLILTEGDLLSAGVDISVQPPGKKIQSLN 1080
 EVK RFK TFEAI+ SFKETF QMFGGGSADL+LTE DLLSAG++ISVQPPGKKIQSLN
 Sbjct: 1021 SEVKARFKVTFEAIQKSFKETFTQMFGGGSADLVLTETDLLSAGIEISVQPPGKKIQSLN 1080

75 Query: 1081 LMSGGEKALSLALLFAIIRVKTIPFVILDEVEAALDEANVKRGDYLNRFDKSSQFIVV 1140
 LMSGGEKALSLALLFAIIRVKTIPFVILDEVEAALDEANVKRGD+LNRFDK SQFIVV
 Sbjct: 1081 LMSGGEKALSLALLFAIIRVKTIPFVILDEVEAALDEANVKRGDFLNRFDKDSQFIVV 1140

80 Query: 1141 THRKGTMMSAADSIYGVTMQESGVSKIVSVKLKEAQMEN 1179
 THRKGTM+AADS+IYG+TMQESGVSKIVSVKLKEAQMEN
 Sbjct: 1141 THRKGTMMSAADSIYGITMQESGVSKIVSVKLKEAQMEN 1179

SEQ ID 1162 (GBS199) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 52 (lane 2; MW 75kDa).

GBS199-GST was purified as shown in Figure 208, lane 3.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 359

A DNA sequence (GBSx0390) was identified in *S.agalactiae* <SEQ ID 1165> which encodes the amino acid sequence <SEQ ID 1166>. This protein is predicted to be ribonuclease III (rnc). Analysis of this protein sequence reveals the following:

```

10 Possible site: 46
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.3372 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9711> which encodes amino acid sequence <SEQ ID 9712> was also identified.

20 The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13466 GB:Z99112 ribonuclease III [Bacillus subtilis]
  Identities = 115/230 (50%), Positives = 154/230 (66%), Gaps = 1/230 (0%)
```

```

25 Query: 13 KKMKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLLNISHNERLEFLGDAVLQLLI 72
          KK+++ + E+ + F N++LL AFTH+SY NEHR NERLEFLGDAVL+L I
          Sbjct: 15 KKVEQFKEFQER-ISVHFQNEKLLYQAFTHSSYVNEHRKKPYEDNERLEFLGDAVLELTI 73
```

```

Query: 73 SQYLFTKYPQKAEGDLSKLRSMIVREESLAGFSRSLCGFDHYIKLGKGEKGSGGRNRDTIL 132
      S++LF KYP +EGDL+KLR+ IV E SL + F + LGKGEE +GGR R +L
      Sbjct: 74 SRFLFAKYPAMSEGDLTKLRAAIVCEPSLVSLAHELSFGDIVLLGKGEEMTGGRKRPALL 133
```

```

Query: 133 GDLFEAFLGALLLDKGVEVVHAFVNKVMIPHVEKGYERVKDYKTSLQELLQSHGDVKID 192
      D+FEAF+GAL LD+G+E V +F+ + P + G + V D+K+ LQE +Q G ++
      Sbjct: 134 ADVFEAFIGALYLDQGLEPVESFLKVYVFPKINDGAFSHVMDFKSQQLQEQYVQRDGKGSLE 193
```

```

35 Query: 193 YQVTNESGPAHAKEFEVTVSVNQENLSQGIGRSKKAAEQAQDAKNALATLQ 242
      Y+++NE GPAH +EFE VS+ E L G GRSKK AEQ AA+ AIA LQ
      Sbjct: 194 YKISNEKGPAHNREFEAIVSLKGEPLGVGNGRSKKEAEQHAAQEALAKLQ 243
```

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1167> which encodes the amino acid sequence <SEQ ID 1168>. Analysis of this protein sequence reveals the following:

```

Possible site: 32
    >>> Seems to have no N-terminal signal sequence

    ----- Final Results -----
    bacterial cytoplasm --- Certainty=0.1414 (Affirmative) < succ>
    bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
    bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

50 An alignment of the GAS and GBS proteins is shown below:

```
Identities = 170/227 (74%), Positives = 192/227 (83%)
```

```
Query: 15 MKELRSKLEKDYGIVFANQELLDTAFTHTSYANEHRLLNISHNERLEFLGDAVLQLLISQ 74
```

-456-

```

      MK+L   L + I F + LL+TAFTHTSYANEHRLLN+SHNERLEFLGDAVLQL+IS+
Sbjct: 1 MKQLEELLSTSFDIQFNDLTLLETAFTHTSYANEHRLNVSHNERLEFLGDAVLQLIISE 60

5     Query: 75 VLFTKYPQKAEGDLSKLRSMIVREESLAGFSRLCGFDHYIKLGKGEEKSGGRNRTDILGD 134
          YLF KYP+K EGD+SKLRSMIVREESLAGFSR C FD YIKLGKGEEKSGGR RDTILGD
Sbjct: 61 YLFAKYPKKTEGDMSKLRSMIVREESLAGFSRCSFDAYIKLGKGEEKSGGRRTDILGD 120

Query: 135 LFEAFLGALLLDKGVEVVHAFVNKMIPHVEKGTYERVKDYKTSLQELLQSHGDVKIDYQ 194
10    Sbjct: 121 LFEAFLGALLLDKGIDAVRRFLKQVMIPQVEKGNNFERVKDYKTCLQEFLQTKGDVAIDYQ 180

Query: 195 VTNESGPAHAKFEVTVSVNQENLSQGIGRSKAAEQDAAKNALATL 241
      V +E GPAHAK+FEV++ VN LS+G+G+SKK AEQDAAKNALA L
Sbjct: 181 VISEKGPAHAKQFEVSIVVNGAVLSKGLGKSKKLAEQDAAKNALAQL 227
15

```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 360

A DNA sequence (GBSx0391) was identified in *S.agalactiae* <SEQ ID 1169> which encodes the amino acid sequence <SEQ ID 1170>. Analysis of this protein sequence reveals the following:

```

Possible site: 43
>>> Seems to have a cleavable N-term signal seq.
INTEGRAL Likelihood = -4.19 Transmembrane 100 - 116 ( 99 - 117)
INTEGRAL Likelihood = -2.44 Transmembrane 81 - 97 ( 81 - 97)

25      ----- Final Results -----
          bacterial membrane --- Certainty=0.2678(Affirmative) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

```

The protein has homology with the following sequences in the GENPEPT database:

```

>GP: CAC12789 GB: AJ279090 hypothetical protein [Staphylococcus
carnosus]
      Identities = 50/114 (43%), Positives = 72/114 (62%)
35      Query: 3 KIFYISLGFISLGIAGIVLPPVPTPLVLLSAFCFSRSSEKFDIWLQRQTKVYKYYAAD 62
          K ++LG I GIG GIV+P++PTTP +LL+A CFSRSS+KF+ WL TK++ Y
      Sbjct: 2 KYVLMTLGLIFAGIGFVGIVVPLLPPTPFLLLAAICFSRSSKKFNRWLVNTKIHDEYVES 61

40      Query: 63 FVESRSTAPARKKSMIWQIYIILMGISIYFAPLMWLKLGLLIGTIVGTYVLFYVV 116
          F + +K ++ +YILMGISI+ +++++ LLV V T VLF V
      Sbjct: 62 FKRDKGFTLKKFKLLTSLYIILMGISIIFIIDNLYIRITLLIMLFVQTVVLFPTFV 115

```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 361

A DNA sequence (GBSx0392) was identified in *S.agalactiae* <SEQ ID 1171> which encodes the amino acid sequence <SEQ ID 1172>. Analysis of this protein sequence reveals the following:

```

50      Possible site: 45
      >>> Seems to have no N-terminal signal sequence

      ----- Final Results -----
          bacterial cytoplasm --- Certainty=0.1908(Affirmative) < succ>
          bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
          bacterial outside --- Certainty=0.0000(Not Clear) < succ>

```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1173> which encodes the amino acid sequence <SEQ ID 1174>. Analysis of this protein sequence reveals the following:

Possible site: 45

5 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1610 (Affirmative) < succ>

10 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 225/269 (83%), Positives = 248/269 (91%)

15 Query: 1 MSEIGFKYSILASGSGTNCFYIETPKRLLIDAGLTGKKVTSLLAEINRKPEDLDAILVT 60
M+E GFKYSILASGSGTNCFY+ETP+KRLLIDAGLTGKK+TSLLAEI+RKPEDLDAIL+T

Sbjct: 1 MNESGFKYSILASGSGTNCFYLETPKRLLIDAGLTGKKITSLLAEIDRKPEDLDAILIT 60

20 Query: 61 HEHSDHIKGVGVLARKYHLDIYANEQTWKVMDERNMLGKVDVSQKHVFGRGKTLTFGDLD 120
HEHSDHIKGVGVLARKYHLDIYANE+TW++MDE NMLGK+D SQKH+F R K LTFGD+D

Sbjct: 61 HEHSDHIKGVGVMARKYHLDIYANEKTWQLMDECNMLGKLDAASQKHIFQRDKVLTFGDVD 120

Query: 121 IESFGVSHDAVPQFYRMMKDDKSFVMLTDTGIVSDRMAGLIEADGYLIESNH DIEILR 180
IESFGVSHDA+DPQFYR+MKD+KSFVMLTDTGIVSDRM G+IENADGYLIESNH DIEILR

25 Sbjct: 121 IESFGVSHDAIDPQFYRIMKDNSFVMLTDTGIVSDRMTGIENADGYLIESNH DIEILR 180

Query: 181 SGSYWPWLKQRILSDKGHLSNEDGSETMIRTIGNRTKHIYLGHLSKENNIKE LAHMTMEN 240
SGSYWP+LKQRILSD GHLSNEDG+ MIR++G TK IYLGHLSKENNIKE LAHMTM N

Sbjct: 181 SGSYWPWLKQRILSDLGHLSNEDGAGAMIRSLGYNTKKIYLGHLSKENNIKE LAHMTMVN 240

30 Query: 241 NLMRADFGVGTDFSVHDTSPDSATPLTRI 269

L AD VGTDF+VHDTSPD+A PLT I

Sbjct: 241 QLAMADLAVGTDFTVHDTSPDTACPLTDI 269

35 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 362

A DNA sequence (GBSx0393) was identified in *S.agalactiae* <SEQ ID 1175> which encodes the amino acid sequence <SEQ ID 1176>. Analysis of this protein sequence reveals the following:

40 Possible site: 26

>>> Seems to have an uncleavable N-term signal seq

INTEGRAL Likelihood ==11.94 Transmembrane 15 - 31 (5 - 34)

----- Final Results -----

45 bacterial membrane --- Certainty=0.5776 (Affirmative) < succ>

bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1177> which encodes the amino acid sequence <SEQ ID 1178>. Analysis of this protein sequence reveals the following:

Possible site: 26

>>> Seems to have a cleavable N-term signal seq.

55 ----- Final Results -----

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>

bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 335/443 (75%), Positives = 392/443 (87%)

```

5      Query: 7  NIRSFEALLFLLVFVAVYFVYLA VRDFKMSKNIRLLNKVRDLIAGNYSDSILIQGDAD 66
          N+ +FELA+L LLVFVA YF++LA VRD++ ++ IR+++ K+RDLI G Y+D I + D +
      Sbjct: 8  NLSTFELAILILLVFVAFYFIHLA VRD YRNARI I RMM SHKIRD LINGRYTDI IDEKADIE 67

10     Query: 67  LVELGESLNLDLSDVFRMAHDNLEQEKNRLASILT YMTDGVLATDRSGKIVMINETAQQF 126
          L+EL + LNDLSDVFR+ H+NL QEKNRLASIL YM+DGVLATDRSGKII+MINETA++Q
      Sbjct: 68  LMELSDQLNLDLSDVFR LTHE NLA QEKNRLASILA YMSDGVLATDRSGKIIIMINETARKQL 127

15     Query: 127  NLAYDEALSMNIVDMLGSGSPYSFQDLVSKTPBEVVLNRDENGEFTLRLRFALNRRESC 186
          NL+ +EAL NI D+L + Y++DLVSKTP V +N R++ GEFV+LR+RFALNRRESC
      Sbjct: 128  NLSKEEALKKNITDLLEGDTSYTYRDLVSKTPVVTVNSRNDMGEFVSLRLRFALNRRESC 187

20     Query: 187  FISGLVAVSHDATEQEKEERERRLFVSNVSH ELRTPLTSVKSYLEALDEGALNEEVAPSF 246
          FISGLV V HD TEQEKEERERRLFVSNVSH ELRTPLTSVKSYLEALDEGAL E++APSF
      Sbjct: 188  FISGLVVVLHDTTEQEKEERERRLFVSNVSH ELRTPLTSVKSYLEALDEGALKEDIAPSF 247

25     Query: 247  IKVSLDETNRMRRMISDLSSLSRIDNEVTHLDVEMTNFTAFMTSILNRFDQIRNQKTVTG 306
          IKVSLDETNRMRRMISDL+LSRIDN+VT L VEMTNFTAF+TSILNRFD ++NQ T TG
      Sbjct: 248  IKVSLDETNRMRRMISDLNLNSRIDNQVTQLAVE MTNFTAFITSILNRFDLVKNQHTGTG 307

30     Query: 307  KVYEIVRDYPLKSIWVEIDNDKMTQVIE NNAIKYSPDGGKITV RMKTTDQLIISIS 366
          KVYEIVRDYP+ S+W+EID DKMTQVI+NIINNA+KYSPDGGKITV ++TT TQ+I+SIS
      Sbjct: 308  KVYEIVRDYPITSVWIEIDNDKMTQVIE NNAIKYSPDGGKITV RMKTTDQLIISIS 367

35     Query: 367  DQGLGIPKKDLPLIFDRFYRVDKARSRKQGGTGLGLSIAKEIVKQHKGFIWAKSEYKGGS 426
          DQGLGIPK DLPLIFDRFYRVDKARSQ QGGTGLGL+IAKEI+KQH GFIWAKS+YKGGS
      Sbjct: 368  DQGLGIPKTDLPLIFDRFYRVDKARSRAQGGTGLGLIAKEIIKQHHGFIWAKSDYKGGS 427

        Query: 427  TFTIVLPYDKDAV TYEE WEDVED 449
          TFTIVLPY+KDA YEEWE+ D
      Sbjct: 428  TFTIVLPYEKDAAIYEE WEE DVD 450

```

A related GBS gene <SEQ ID 8561> and protein <SEQ ID 8562> were also identified. Analysis of this protein sequence reveals the following:

```

40     Lipop: Possible site: -1 Crend: 8
      McG: Discrim Score: 8.59
      GvH: Signal Score (-7.5): -3.38
          Possible site: 26
      >>> Seems to have an uncleavable N-term signal seq
      ALOM program count: 1 value: -11.94 threshold: 0.0
      INTEGRAL Likelihood = -11.94 Transmembrane 15 - 31 ( 5 - 34)
      PERIPHERAL Likelihood = 8.27 178
      modified ALOM score: 2.89

```

```

50     *** Reasoning Step: 3
      ----- Final Results -----
          bacterial membrane --- Certainty=0.5776 (Affirmative) < succ>
          bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
          bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

```

55 The protein has homology with the following sequences in the databases:

67.5/83.5% over 439aa

Streptococcus pneumoniae

GP|5830524| histidine kinase Insert characterized

```

60     ORF01458(331 - 1647 of 1947)
     GP|5830524|emb|CAB54569.1||AJ006392(10 - 449 of 449) histidine kinase {Streptococcus
     pneumoniae}
     %Match = 45.6

```

-459-

%Identity = 67.5 %Similarity = 83.4
 Matches = 297 Mismatches = 70 Conservative Sub.s = 70

5	126	156	186	216	246	276	306	336
	ITSPFSDTYRTSHDTRFIGNSLGI*LFWRCPPS*CDGETFT*KD*RYSWSSRIYFDSTWCRXIT*SLMNNSAANIRSFE							
								MLDLLKQTIFT
								10
10	366	396	426	450	480	510	540	570
	LALLFLVFAVYFVYLAVRDFKMSKNIRL--LNWKVRDLIAGNYSDSILIQGDADLVELGESLNDLSDVFRMAHDNLEQ							
	:: : : ::	:	:	:: : : : :	:: :	:: : : :	:: : : :	:: : : :
	RDFIFILILLGFILVVTLLLLENRRDNIQLKQVNQKVLDLIAQDYSKVLMQGGSEITNTITNNNLNDLSEVIRLTQENLEQ							
	30	40	50	60	70	80	90	
15	600	630	660	690	720	750	780	810
	EKNRLASILTYMTDGVLATDRSGKIVMINETAQQQFNLAYDEALSMNIVDMLGSGSPYSFQDLVSKTPEVVLNRDENG							
			:	: :	:: :	::	: : :	
	ESKRILNSILFYMTDGVLATNRGQIIMINDTAKQLGLVKEDVLNRSLIELLKIEENYELRDLITQSPELLDSQDINGE							
	110	120	130	140	150	160	170	
20	840	870	900	930	960	990	1020	1050
	FVTLRIRFALNRRESGFISGLVAVSHDATEQEKEERERRLFVSNVSHELRPLTSVKSYLEALDEGALNEEVAPSFIKV							
	:: : :		:		:		:	:
	YLNLRVRFALIRRESGFISGLVAVLHDTTEQEKEERERRLFVSNVSHELRPLTSVKSYLEALDEGALCETVAPDFIKVS							
	190	200	210	220	230	240	250	
25	1080	1110	1140	1170	1200	1230	1260	1290
	LDETNRMRRMISDLSSLSRIDNEVTHLDVEMTNFTAFMTSILNRFQDQIRNQKTVTGKVYEIVRDYPLKSIWVEIDTDKMT							
		:	: :	:	:		:	:
	LDETNRMRRMVTDLHLRSRIDNATSHLDVELNFTAFITFILNRFDKMKQQ--EKEKKYELVRDYPINSIWMIEIDTDKMT							
	270	280	290	300	310	320		
30	1320	1350	1380	1410	1440	1470	1500	1530
	QVIDNILNNAVKYSPDGGKITVNLRTTKTOMILSISDQGLGIPKKDLPLIFDRFYRVDKARSRKQGGTGLGLSIAKEIVK							
	: :	: :	:: :	:	:	:	:	:
	QVVDNILNNAIKYSPDGGKITVRMKTEDQMLISISDHGLGIPKQDLPRIFDRFYRVDRARSRAQGGTGLGLSIAKEIIK							
	340	350	360	370	380	390	400	
35	1560	1590	1620	1647	1677	1707	1737	1767
	QHKGFIWAKSEYGKGSTFTIVLPYDKDAVTYEEWED-VED*NMSEIFGKYSILASGSTGNCFYIETPKRLLIDAGLTGK							
		:						
	QHKGFIWAKSEYGKGSTFTIVLPYDKDAVKEEWEDEVED							
	420	430	440					
40								
45								

SEQ ID 1176 (GBS41) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 7; MW 50kDa), in Figure 168 (lane 2-4; MW 65kDa – thioredoxin fusion) and in Figure 238 (lane 4; MW 65kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 13 (lane 7; MW 75kDa).

Purified Thio-GBS41-His is shown in Figure 244, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 363

A DNA sequence (GBSx0394) was identified in *S.agalactiae* <SEQ ID 1179> which encodes the amino acid sequence <SEQ ID 1180>. This protein is predicted to be VicR protein (regX3). Analysis of this protein sequence reveals the following:

Possible site: 60
 >>> Seems to have no N-terminal signal sequence

-460-

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2754 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1181> which encodes the amino acid sequence <SEQ ID 1182>. Analysis of this protein sequence reveals the following:

10 Possible site: 60

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

15 bacterial cytoplasm --- Certainty=0.2754 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 205/236 (86%), Positives = 221/236 (92%)

20 Query: 1 MKKILIVDDEKPISDIKFNLTKEGYETATAFDGREALVQYAEFQPDLIILDLMLPFLDG 60
 MKKILIVDDEKPISDIKFNLTKEGY+ TAFDGREA+ + E + PDLIILDLMLPFLDG

Sbjct: 1 MKKILIVDDEKPISDIKFNLTKEGYDIVTAFDGREALTIFEEEKPDLIILDLMLPFLDG 60

25 Query: 61 LEVAKEVRKTSHPPIIMLSAKDSEFDKVIGLEIGADDYVTKPFNSNELLARVKAHLRTE 120
 LEVAK+E+RKTSH+PIIMLSAKDSEFDKVIGLEIGADDYVTKPFNSNELLARVKAHLRTE

Sbjct: 61 LEVAKEIRKTKTSHVPIIMLSAKDSEFDKVIGLEIGADDYVTKPFNSNELLARVKAHLRTE 120

Query: 121 NIETAVAAEESAQNASSDITIGELQILPDAFIKKRGEIELTHREFELLHHLATHIGQVM 180
 IETAVAAE+E+A + + ++TIG LQILPDAF+AKK G+E+ELTHREFELLHHLA H+GQVM

30 Sbjct: 121 TIETAVAAEENASSGTQELTIGNLQILPDAFVAKKHGQEVELTHREFELLHHHLANHMGQVM 180

Query: 181 TREHLLLETWVGYDYFGDVRTVDVTVRRLREKIEDTPGRPEYLTRRGVGYYMKSYE 236
 TREHLLLE VWGVDYFGDVRTVDVTVRRLREKIEDTP RPEYLTRRGVGYYMKSYE+

35 Sbjct: 181 TREHLLLEIVWVGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYLTRRGVGYYMKSYD 236

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 364

A DNA sequence (GBSx0395) was identified in *S.agalactiae* <SEQ ID 1183> which encodes the amino acid sequence <SEQ ID 1184>. This protein is predicted to be amino acid ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

40 Possible site: 43

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

45 bacterial cytoplasm --- Certainty=0.3791 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

50 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14701 GB:Z99118 glutamine ABC transporter (ATP-binding protein) [Bacillus subtilis]
 Identities = 149/244 (61%), Positives = 200/244 (81%), Gaps = 2/244 (0%)

55 Query: 3 LISYKNVNKYYG DYHALQINLEIEPGQVVVLGPGSGKSTLIRTMNALESIDDSL VV 62
 +I++ +N VNK+ YGD+H L+QINL+IE G+VVV++GPGSGSKSTL+R +N LESI++G L V
 Sbjct: 1 MITFQNVNKH YGDHFV LKQINLQIEKGEV VVII GPGSGSKSTLRC INRLESINEGV LTV 60

-461-

Query: 63 NGHELIANSKELVNLKEVGMVFQHFNLYPHKTVLENITLAPIKVVKQSKKEAMEIAEK 122
 NG + N ++ +R+ +GMVFQHF+LYPHKTVL+NI LAP+KVL+QS ++A E A
 Sbjct: 61 NGTAI-NDRKTDINQRQNIGMVFQHPHLYPHKTVLQNMILAPVKVLRQSPEQAKETARY 119

5 Query: 123 YLKVFVNMRWERKDSYPSMLSQQGKQRIAIARGLAMHPKLLLDEPTSALDPETIGDVLSV 182
 YL+ V + ++ D+YPS LSGGQ+QR+AIARGLAM P+++LFDEPTSALDPE IG+VL VM
 Sbjct: 120 YLEKVGIPDKADAYPSQLSGGQQQRVAIARGLAMKPEVMLFDEPTSALDPEMIGEVLDVM 179

10 Query: 183 QKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFFDNPREPRAKQFLSN 242
 + LA +GM MVVVTHEMGFA+EVADRI+F+ +G+IL + +F+ NP+E RA+ FLS
 Sbjct: 180 KTLAKEGMTMVVVTHEMGFAKEVADRIIVIDEKGKILEEAVPA-EFYANPKEERARLFLSR 238

15 Query: 243 IINH 246
 I+NH
 Sbjct: 239 ILNH 242

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1185> which encodes the amino acid sequence <SEQ ID 1186>. Analysis of this protein sequence reveals the following:

20 Possible site: 51
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3763(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 131/243 (53%), Positives = 179/243 (72%), Gaps = 2/243 (0%)

30 Query: 2 SLISYKNVNKYYGDYHALRQINLEIEPGQVVVLLGPSGSKSTLIRTMNALESIDDSL 61
 ++IS K+++KYYG L+ I+L+I PG+VVV++GPSGSKSTL+RTMN LE G +
 Sbjct: 5 AIISIKDLHKYYGHNEVLKGIDLDIMPGEVVVIIGPSGSGKSTLLRTMNLLEVPTKGQIR 64

35 Query: 62 VNGHELANSKELVNLKEVGMVFQHFNLYPHKTVLENITLAPIVKVKQSKKEAMEIAE 121
 G ++ + ++ ++R+++GMVFQ FNL+P+ T+LENITL+PIK +K EA + A
 Sbjct: 65 FEGIDITD-KKNDIFSMREKMGMVQQFNLPNMTILENTLSPIKTKGMAKEADKTAL 123

40 Query: 122 KYLKVFVNMRWERKDSYPSMLSQQGKQRIAIARGLAMHPKLLLDEPTSALDPETIGDVLSV 181
 L V + E+ +YP+ LSGGQ+QRIAIARGLAM P +LLFDEPTSALDPE +G+VL+V
 Sbjct: 124 SLLDKVGLSEKAKAYPASLSGGQQQRVAIARGLAMDPDVLLFDEPTSALDPEMVGEVLAV 183

45 Query: 182 MQKLANDGMNMVVVTHEMGFAREVADRIIFMADGEILVDTTDVQDFFDNPREPRAKQFLS 241
 MQ LA GM MV+VTHEMGFA+EVADR++FM DG ++V+ FD +E R K FLS
 Sbjct: 184 MQDLAKSGMTMIVTHEMGFAKEVADRMFM-DGGVIVEEGSPNQLFDLTKEERTKDFLS 242

Query: 242 NII 244
 ++
 Sbjct: 243 RVL 245

50 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 365

A DNA sequence (GBSx0396) was identified in *S.agalactiae* <SEQ ID 1187> which encodes the amino acid sequence <SEQ ID 1188>. This protein is predicted to be glutamine-binding. Analysis of this protein sequence reveals the following:

Possible site: 27
 >>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

-462-

bacterial outside --- Certainty=0.3000 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

5 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB73178 GB:AL139076 probable ABC-type amino-acid transporter
 periplasmic solute-binding protein [Campylobacter
 jejuni]
 Identities = 99/240 (41%), Positives = 141/240 (58%), Gaps = 3/240 (1%)
 10 Query: 1 MLRRKRRLTFYLLSCIFIFLLFYPNSTSANQLSEIKKSGVLKVGVKQDVNPFGYYNAETNQ 60
 M+ RK L + + + F + + + L IK G L VGVK DVP++ + T +
 Sbjct: 1 MVFRKSLLKLAVFALGACVAFSNANAAEGKLESIKSKGQILVGVKNDVPHYALLDQATGE 60
 15 Query: 61 YEGMEIDIakkIAkSL---GVKPVFVPTTAQTRPLMDNGQIDILIATYTITPERKANYN 117
 +G E+D+AK +AKS+ K V A+TR PL+DNG +D +IAT+TITPERK YN
 Sbjct: 61 IKGFEVVDVAKLLAKSILGDDKKIKLVAVNAKTRGPLLNGSVDAVIATTITPERKRIYN 120
 20 Query: 118 ISKAYYHDEIGFLVLRKNSHIKTIKELDGKHIGVAQGATTKVNLKYAKEHKLKFSYAQLG 177
 S+ YY D IG LV K K++ ++ G +IGVAQ ATT K + + AK+ + ++++
 Sbjct: 121 FSEPYYQDAIGLLVLKEKKYKSLADMKGANIGVAQAATTKAIGEAAKIGIDVKFSEFP 180
 25 Query: 178 SFPELAISLYANRIDAFAVDKSILSGYLSPTHITKEGFTNQEQYGIATSKQDKVLIPYVN 237
 +P + + L A R+DAFSVDKSIL GY+ + IL + F Q YGI T K D YY+
 Sbjct: 181 DYPSTKAALDAKRVDAFSDVKSILLGYVDDKSEILPDSFEPQSYGIVTKKDDPAFAKYVD 240

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1189> which encodes the amino acid sequence <SEQ ID 1190>. Analysis of this protein sequence reveals the following:

Possible site: 30
 30 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -6.16 Transmembrane 17 - 33 (15 - 35)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3463 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

A related sequence was also identified in GAS <SEQ ID 9097> which encodes the amino acid sequence <SEQ ID 9098>. Analysis of this protein sequence reveals the following:

40 >>> May be a lipoprotein
 ----- Final Results -----
 bacterial membrane --- Certainty= 0.000 (Not Clear) < succ>
 bacterial outside --- Certainty= 0.000 (Not Clear) < succ>
 45 bacterial cytoplasm --- Certainty= 0.000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 66/251 (26%), Positives = 111/251 (43%), Gaps = 27/251 (10%)

50 Query: 23 PNSTSANQLSEIKKSGVLKVGVKQDVNPFGYYNAETNQYEGMEIDIakkIAkSLGVKPVF 82
 P+ + + IK+ GVLKV +YN + N+ G E+D+ K+I K L +K F
 Sbjct: 34 PHQSQKSSWDTIKEKGVLKVATPGTYQPTSFYN-DNNELVGYEVDMVKEIGKRLNIKVKF 92
 55 Query: 83 VPTTAQTRPLMDNGQIDILIATYTITPERKANYNISKAYYHDEIGFLVR---KNSHIK 138
 V T +D+G++DI + + ITP+R+ YNIS Y + G +VR N K
 Sbjct: 93 VETGFDQAFTSVDSGRVDISLNNFDITPKRQKKYNISTPYKYGVGGMIVRADGSSIakk 152
 Query: 139 TIKELDGKHIGVAQGATTKVNLKYAKEHKLKFSYAQLGSFPELAISLYANRI----- 191
 + + GK A G +K A+L ++ + +Y N +
 60 Sbjct: 153 DLSDWKGGKKAAGASGTEYMKVAQKQG-----AELVTYDNVTGDVYLNDVANGRTDF 203
 Query: 192 --DAFSVDKSILSGYLSPTHITKE---GFNTQEQYGIATSKQDKVLIPYVNKLVSWEK 245

+ + K + L S + + + + N E G I + K + D L ++ + + K
 Sbjct: 204 IPNDYPAQKLFVDYMLSQNPNLNVKMSDVQYNPTEQQGIVMNKKDDSLKKKIDAVIKDMIK 263

5 Query: 246 DGSLKHIIYQKF 256
 DGSLK I + +
 Sbjct: 264 DGSLKKISETY 274

SEQ ID 1188 (GBS136) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 29 (lane 5; MW 29.9kDa).

10 The GBS136-His fusion product was purified (Figure 200, lane 6) and used to immunise mice. The resulting antiserum was used for FACS (Figure 284), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

15 Example 366

A DNA sequence (GBSx0397) was identified in *S.agalactiae* <SEQ ID 1191> which encodes the amino acid sequence <SEQ ID 1192>. This protein is predicted to be integral membrane. Analysis of this protein sequence reveals the following:

20 Possible site: 55
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -9.34 Transmembrane 32 - 48 (27 - 55)
 INTEGRAL Likelihood = -5.04 Transmembrane 200 - 216 (196 - 219)
 INTEGRAL Likelihood = -3.13 Transmembrane 93 - 109 (93 - 113)
 INTEGRAL Likelihood = -2.02 Transmembrane 74 - 90 (74 - 92)
 25 ----- Final Results -----
 bacterial membrane --- Certainty=0.4736(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB73177 GB:AL139076 putative ABC-type amino-acid transporter
 permease protein [Campylobacter jejuni]
 35 Identities = 112/226 (49%), Positives = 160/226 (70%), Gaps = 3/226 (1%)
 Query: 5 NISPFAISRWGAGFFNHFDLFFKGFLYTLGISFGALLLALILGILSGGLSTS SKVGKLIS 64
 +ISPFA+ ++ ++ D F GF+YTL +S ALL+A I G + G ++TS+ K+ + +
 Sbjct: 25 SISPFAVWKFLDALDNKDAFINGFIYTLLEVSILALLIATIFGTIGGVMATS RFKII RAYT 84
 40 Query: 65 RIYVEVFQNTPLLVQMVFVYYGLAIISNGHVMISAFFTAVL CVGLYHGAYISEVIRSGIE 124
 RIYVE+FQN PL++Q+ F++Y L ++ + + F VL VG YHGAY+SEV+RSGI
 Sbjct: 85 RIYVELFQNVPLVIQIFFLFYALPVLG---IRLDIFTIGVLGVGAYHGAYVSEVVRSGIL 141
 45 Query: 125 AVPKGGQTEAALAQQGFTANQTMQLIILPQAVRTILPPMTNQVVNLKNTSTVAIISGADIM 184
 AVP+GQ EA+ +QGFT Q M+ II+PQ +R ILPPMTNQ+VNLIKNTS + I+ GA++M
 Sbjct: 142 AVPRGQFEASASQGFTYIQQMRYII VPQTIRIILPPMTNQMVNLKNTSVLLIVGGAELM 201
 Query: 185 FVAKAWAYDTTNYIPAFAGAAIFYFVICFPLASWARKQEELNKTY 230
 A ++A D NY PA+ AA+ YF+IC+PLA +A+ E KK +
 50 Sbjct: 202 HSADSYAADYGNYAPAYI FAAVLYFIICYPLAYFAKAYENKLKKAH 247

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1193> which encodes the amino acid sequence <SEQ ID 1194>. Analysis of this protein sequence reveals the following:

55 Possible site: 28
 >>> Seems to have a cleavable N-term signal seq.

-464-

INTEGRAL Likelihood = -6.26 Transmembrane 307 - 323 (303 - 327)
 INTEGRAL Likelihood = -5.89 Transmembrane 485 - 501 (479 - 502)
 INTEGRAL Likelihood = -1.12 Transmembrane 375 - 391 (375 - 391)

5 ----- Final Results -----

bacterial membrane --- Certainty=0.3506(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

10 The protein has homology with the following sequences in the databases:

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein
 [Synechocystis sp.]
 Identities = 146/532 (27%), Positives = 244/532 (45%), Gaps = 59/532 (11%)

15 Query: 6 YMKKLILSCLVALALLFGGMSRAQANQYLRVGMEAAYAPFNWTQDDASNGAVPIEGTSQY 65
 Y L L L+A+A+ + Q + V E + PF T E T Q
 Sbjct: 16 YYLLALGVLLATAIPLLPAFSQVSRQTIIVATEPTFPPEMTD-----EATGQL 65

20 Query: 66 ANGYDVQVAKKVKAMNKELLVVKTSWTGLIPALTSGKIDMIAAGMSPTKERRNEISFSN 125
 G+DV + + + A + + + G+IPAL S + + ++ T ER +SFS+
 Sbjct: 66 T-GFDVDLIQAIGEAAQVITVDIQGYPFDGIIPALQSNTVGAAISAITITPERAQSVSFSS 124

25 Query: 126 SSYTSQPVLVVTANGKYADATSLKDFTSGAKVTAQQGVWHVNLLTQLKGAKLQTPMGDFSQ 185
 + S VL + +LKD G ++ G + T + GAK+ T +
 Sbjct: 125 PYFKS--VLAIAVQDGNDTIKNLKDLEGKRLAVAIGTTGAMVATNVPGAKV-TNFDSITS 181

30 Query: 186 MRQALTSGVIDAYISERPEAMTAEEAADSRLKMITLKKGFAVAESDAIAAVGMKKNDRMA 245
 Q L +G DA I++RP + A D+ L+ + + +E IA+ + +
 Sbjct: 182 ALQELVGNADAVINDRPVLLYA-IKDALGRNVKISADVG-SEYYGIAMPLAPPGE--- 236

Query: 246 TVNQVLEGFSQTDRMALMDDMVTKQPVEKKAEDAKASFQFLGQMWAIFKGN----- 294
 +NQ E +Q +++++ EK + FL + G
 Sbjct: 237 -INQTREVLNQ-GLFQIIENGTYNAIYEKFGEKNPPFLPLVAPSILVGKVGTQSLTERS 294

35 Query: 295 -----WKQFLRGTGMTLLISMVGITITGLFIGLLIGIFRTAPKAKHKVAALGQK 342
 ++ +G+ +T+L++ GL G + I + K
 Sbjct: 295 QANPNDNFLTLEFRNLFKGSILTVLLTAFSVFFGLIGGTGVAIALISDI-----K 344

40 Query: 343 LFGWLLTIYIEIFRGTPMIVQSMVIYYGTAQAF----GISIDRTLAAIFIIVSINTGAYM 397
 + IY+E FRGTPM+VQ +IY+G F GI+IDR AAI +S+N AY+
 Sbjct: 345 PLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEIGLGITIDRFPAIIALSLSNVAAYL 404

45 Query: 398 SEIVRGGIFAVDKQFKAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTSVL 457
 +EI+RGGI ++D+GQ++A +LG + QTMM+++ PQ R ILP GNEF+ IKDTS+
 Sbjct: 405 AEIIRGGIQSIDQGQWEACESLGMSPWQTMKEVIFPQAFRRILPPLGNEFITLIKDTSLT 464

50 Query: 458 NVISVVVELYFSGNTVATQTYQYFQTFITIAYFVLTFVTRILRYIERRFD 509
 VI EL+ G + TY+ F+ + A++Y +LT + + +++E D
 Sbjct: 465 AVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTTISSFVFKWLENYMD 516

An alignment of the GAS and GBS proteins is shown below:

Identities = 82/210 (39%), Positives = 113/210 (53%), Gaps = 12/210 (5%)

55 Query: 14 WGAFNHFDLFFKGFLYTLGIGALLLALILGILSGGLSTS---KSKVGKL-----I 63
 W F ++ F +G TL IS + L +G+L G T+ K KV L +
 Sbjct: 288 WAIFKGNWKQFLRGTGMTLLISMVGITITGLFIGLLIGIFRTAPKAKHKVAALGQKLFQWL 347

60 Query: 64 SRIYVEVFQNTPLLVQMFVYYGLAIISNGHVMISAFFTAVLCVGLYHGAYISEVIRSGI 123
 IY+E+F+ TP++VQ + YYG A + I A+ V + GAY+SE++R GI
 Sbjct: 348 LTIYIEIFRGTPMIVQSMVIYYGTAQAFG--ISIDRTLAAIFIIVSINTGAYMSEIVRGGI 405

65 Query: 124 EAVPKGQTEAAALAQGFTANQTMQLIILPQAVRTILPPMTNQVNNLIKNTSTVAIISGADI 183
 AV KGQ +AA A GFT QTMM+ I+LPQ VR ILP N+ V IK+TS + +IS ++
 Sbjct: 406 FAVDKGQFKAAATALGFTHGQTMRKIVLPQVVRNILPATGNEFVINIKDTSVLNVISVVEL 465

Query: 184 MFVAKAWAYDTTMNYIPAFAGAAIFYFVICF 213

-465-

F A T Y F AI YFV+ F
 Sbjct: 466 YFSGNTVATQTYQYFQTFTIIAIYFVLTF 495

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 5 vaccines or diagnostics.

Example 367

A DNA sequence (GBSx0398) was identified in *S.agalactiae* <SEQ ID 1195> which encodes the amino acid sequence <SEQ ID 1196>. This protein is predicted to be amino acid ABC transporter, permease protein. Analysis of this protein sequence reveals the following:

10 Possible site: 39
 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -6.95 Transmembrane 25 - 41 (16 - 42)
 INTEGRAL Likelihood = -3.61 Transmembrane 66 - 82 (65 - 86)
 15 INTEGRAL Likelihood = -2.44 Transmembrane 184 - 200 (182 - 201)
 INTEGRAL Likelihood = -0.59 Transmembrane 119 - 135 (119 - 135)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3781(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 20 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAB14704 GB:Z99118 glutamine ABC transporter (integral membrane
 25 protein) [Bacillus subtilis]
 Identities = 84/206 (40%), Positives = 129/206 (61%), Gaps = 6/206 (2%)
 Query: 10 ILFLLQGFGLTLYISFISIILLSMFFGTLLAIMRNSKNPIWKLIASIYIEFVRNVPNLLWI 69
 + FL GF +TLY++FISI+LS FFG + +R +K P+ + ++ +E +RN+P LL I
 Sbjct: 12 LAFLWDGFLVTLYVAFISIILSFFFGLIAGTLRYAKVPVLSQLIAVLTETIRNLPLLLII 71
 30 Query: 70 FIIIFLVF----QMKSVSAGITSFTIFTSAALAEIIRGGLNGVDKGQTEAGLSQGFTYI 124
 F F +++ +A IT+ TIF SA L+EIIR GL +DKGQ EA S G +Y Q
 Sbjct: 72 FFTFFALPEIGIKLEITAAITALTIFESAMLSEIIRSGLKSIDKGQIEAARSSGLSYTQ 131
 35 Query: 125 VFIIIIIPQAQRKMLPAIIISQFVTVIKDTSLLYS VIAI QEI FGKSQILMGRYFEAGQVFT 184
 I+ PQA R+M+P I+SQF+--+KDTSL VIA+ E+ +OI+ G+ + F
 Sbjct: 132 TLFFIVMPQALRRMVPPIVSQFISLLKDTSLAV-VIALPELIHNAQIINGQSADGSYFP 190
 Query: 185 LYAIITAVYFITNFISSFSRKLSKR 210
 40 ++ + +YF N+ +S +R+L R
 Sbjct: 191 IFLLAALMYFAVNYSLSLAARRLEVR 216

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1197> which encodes the amino acid sequence <SEQ ID 1198>. Analysis of this protein sequence reveals the following:

45 Possible site: 20
 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood = -10.51 Transmembrane 529 - 545 (517 - 551)
 INTEGRAL Likelihood = -10.30 Transmembrane 697 - 713 (693 - 719)
 50 INTEGRAL Likelihood = -4.41 Transmembrane 560 - 576 (555 - 585)
 INTEGRAL Likelihood = -0.32 Transmembrane 662 - 678 (662 - 678)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.5203(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 55 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:BAA17584 GB:D90907 glutamine-binding periplasmic protein

-466-

[Synechocystis sp.]

Identities = 153/475 (32%), Positives = 251/475 (52%), Gaps = 27/475 (5%)

5 Query: 273 IVSDSSFAPFEFQN-GKGKYVGIDIELIKAIAKQQGFKIEIANPGFDAALNAVQSSQADG 331
 + ++ +F PFE + G+ G D++LI+AI + ++I FD + A+QS+

Sbjct: 46 VATEPTFPPEMTEATGQLTGFVDLQIQAIGEAAQVTVDIQCQYGPFDGIIPALQSNTVGA 105

10 Query: 332 VIAGATITDARKAIFDFSDPYYTTSNIIILAVKAGKN-IKNYEDLDRKTVGAKNGTSSYWL 390
 I+ TIT R FS PY+ S + AV+ G + IKN +DL+ K + GT+ + +

Sbjct: 106 AISAITITPERAQSFSFSSPYFKSVLIAIAVQDGNDTIKNLKDLEGKRLAVAIGTTG-AMV 164

15 Query: 391 KENAPKYGYNVKAFDDGSSMYDSLSNSGSVDAIMDDEAVLKYAISQG--RRFETPLEGIST 448
 N P G V FD +S L +G+ DA+++D VL YAI R + + S

Sbjct: 165 ATNVP--GAKVTNFDSITSALQELVNGNADAVINDRPVLLYAIKDAGLRNVKISADVGSE 222

20 Query: 449 GEVGFAVKKGTNPCLI--EMFNNGLAALKSGQYDDIIDKYLDSKKA----ATPSEKG 500
 G A+ E+ E+ N GL + ++G Y+ I +K+ K PS G

Sbjct: 223 DYYGIAMPLAPPGEINQTREVLNQGLFQIIENGTYNAIYEKWFGEKNPPFLPLVAPSLVG 282

25 Query: 501 -----ADESTISGLLSNNYKQLLAGLGTTLSSLTLISFAIAIIIGIIFGMMAVSP 549
 + + L ++ L G T+ LT S +I G + + S

Sbjct: 283 KVGTAQSLTERSQANPNDNFLITLFRNLFKGSILTVLLTAFSVFFGLIGGTGVAILISD 342

30 Query: 550 TKSLRLISTVFVDVVRGIPLMIVAAFIFWGVPNLIESMTGHQSPINDFLAATIALSLNCG 609
 K L+LI ++V+ RG P+++ I++G+P L + + G I+ F AA IALSLN

Sbjct: 343 IKPLQLIFRIYVEFFRGTPMLVQLFIIYFGLPALFKEI-GLGITIDRFPAAIIALSLNVA 401

Query: 610 AYIAEIIRGGIEAVPAGQMEASRSRSLGLSYGTTMRKVILPQAVKMLLPNFINQFVISLKD 669

AY+AEI+RGGI+++ GQ EA SLG+S TM++VI PQA + +LP N+F+ +KDT

Sbjct: 402 AYLAEIIRGGIQSIDQGQWEACESLGMSPWQMKEVIFPQAFRRILPPLGNFETLIKDT 461

Query: 670 TIVSAIGLVELFQTGKIIIARNYQSFRMYAILAIIYLIIMIILLTRILAKRLEKRLN 724

++ + IG ELF+ G++I+A Y++F +Y +A++YL++ + + + K LE ++

Sbjct: 462 SLTAVIGFQELFREGQLIVATTYRAFEVYIAVALVYLLLTISSFVFKLENYMD 516

Identities = 68/247 (27%), Positives = 106/247 (42%), Gaps = 11/247 (4%)

35 Query: 7 VLLLAIMSIFLTCNIASAETIAIASDTAYAPFEFKD--SDQIYKGIDVDIINEVAKRQSW 64
 VLL + + + S +TI + ++ + PFE D + Q+ G DVD+I + +

Sbjct: 24 VLLAIAIPLLPAFSQVSROTIIIVATEPTFPPEMTEATGQL-TGFVDLQIQAIGEAAQV 82

40 Query: 65 DFSMSFPFGDAAVNAVQSGQASALMAGTTITNARKVFHFSEPYDTKIVIATRKAN-AI 123

+ FD + A+QS A ++ TIT R + FS PY+ + + IA + N I

Sbjct: 83 TVDIQGYPFDGIIPALQSNTVGAIAISAITITPERAQSFSFSSPYFKSVLIAIAVQDGNDTI 142

45 Query: 124 KKYSIDLKGKTVGVKNGTAAQAFLNYYKKKYDYTAKTFDTGDLMYNSLSAGSIAAVMDDEA 183

K DL+GK + V GT N V FD+ L G+ AV++D

Sbjct: 143 KNLKDLEGKRLAVAIGTTGAMVATNVP---GAKVTNFDSITSALQELVNGNADAVINDRP 199

50 Query: 184 VIQYAI-----QNQDIAINMKGEPIGSFGFAVKKGSGYDYLVNDFTNTALKAMKADGTQAI 239

V+ YAI +N I+ ++ E + + + N L + +GTY A

Sbjct: 200 VLLYAIKDAGLRNVKISADVGSEDYYGIAMPLAPPGEINQTREVLNQGLFQIIENGTYNA 259

Query: 240 IMTKWLG 246

I KW G

Sbjct: 260 IYEKWF 266

An alignment of the GAS and GBS proteins is shown below:

Identities = 68/210 (32%), Positives = 113/210 (53%), Gaps = 16/210 (7%)

60 Query: 13 LLQGFGLTLYISFISILLSMFFGTLLAIMRNSKNPIWKLIASIYIEFVRNVPNLLWIFI 72
 LL G G TL ++ IS +++ G + +M S +LI+++++ VR +P ++ I

Sbjct: 517 LLAGLGTLTSLTLISFAIAIIIGIIFGMMAVSPTKSLRLISTVFVDVVRGIPLMIVAAFI 576

65 Query: 73 F-----L VFQMKSVSAGITSFTIFT-----SAALAEIIRGGLNGVDKGQTEAGLSQGF 120
 F L+ M + I F T A +AEI+RGG+ V GQ EA S G

Sbjct: 577 FWGVPNLIESMTGHQSPINDFLAATIALSLNGGAYIAEIVRGGIEAVPAGQMEASRSLGL 636

-467-

Query: 121 TYLQVFIIIIIFPQAFRKMLPAIISQFVTVIKDTSLLYSVIAIQEIFGKSQILMGRYFEAG 180
 +Y +I PQA + MLP I+QFV +KDT++ S I + E+F +I++ R +
 Sbjct: 637 SYGTTMRKVILPQAVKLMLPNFINQFVISLKDITIV- SAIGLVELFQTGKIIIARNY-- 692

5 Query: 181 QVFTLYAIITAVYFITNFISSFSRKLSKR 210
 Q F +YAI+ +Y I +++ +++L KR
 Sbjct: 693 QSFRMYAILAITYLIMIILLTRLAKRLEKR 722

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for
 10 vaccines or diagnostics.

Example 368

A DNA sequence (GBSx0399) was identified in *S.agalactiae* <SEQ ID 1199> which encodes the amino acid sequence <SEQ ID 1200>. Analysis of this protein sequence reveals the following:

Possible site: 39
 15 >>> Seems to have an uncleavable N-term signal seq
 INTEGRAL Likelihood = -12.21 Transmembrane 7 - 23 (1 - 30)
 ----- Final Results -----
 20 bacterial membrane --- Certainty=0.5883 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP: BAB04094 GB: AP001508 unknown conserved protein in *B. subtilis*
 25 [Bacillus halodurans]
 Identities = 43/157 (27%), Positives = 83/157 (52%), Gaps = 9/157 (5%)
 Query: 26 YQSQFQKTTNQALAIAYDAKVAKK--DVIHQKIDKEFENFRGSYEIEFNTKSAEYSYHV 83
 +Q++ N+ L +A ++ + + + +K+ +N R YEIE EY + +
 30 Sbjct: 38 HQAESVSADNEGLTLAEASDIALERAGNGVVTEAEKDRDNGRVVYEIEVKNDDDEYDFKI 97
 Query: 84 DVKTGQILERDMNNNGFSKSTSQSSSSSQKSHKISQEEAKKIAFKDANIEESEVSNLKI 143
 D +TG+IL+ + SK SSS ++ IS +EAK+IA K+ + ++ +++++
 35 Sbjct: 98 DQQTGEILKEKQEQRKGSKPREGHSSSKGSEA-VISMDEAKEIALKEVS---GKIDDIEL 153
 Query: 144 KEEIENGKSVYDIDF-VDLKNKNEVDYQIDAETGKII 179
 E ENG VY+++ D + ++V +DA TG ++
 Sbjct: 154 --ERENGSLVYEVIEESDHYDDDVTVYVDAWTGNVL 188

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1201> which encodes the amino acid sequence <SEQ ID 1202>. Analysis of this protein sequence reveals the following:

Possible site: 57
 45 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -5.15 Transmembrane 42 - 58 (41 - 60)
 ----- Final Results -----
 bacterial membrane --- Certainty=0.3060 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 50 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

Identities = 37/96 (38%), Positives = 63/96 (65%), Gaps = 5/96 (5%)
 55 Query: 94 DMDDNNGFSKSTSOSSSSSQKSHKISQEEAKKIAFKDANIEESEVSNLKIKEEIENGKSV 153
 DMD+ +Q +S + K K+S+++AK IA KDA++ E++ L + ++ E+GK+V
 Sbjct: 59 DMDDDKD-DHMDNQPKTSQTSKKVVLSEDKAALKDASVTEADAQMLSVTQDNEDGKAV 117

-469-

762	792	822	852	882	912	942	972
EEIENGKSVYDIDFVLKNKNEVDYQIDAETGKIIERSRDHMND*FK*DIKKRRSKRPSF*LLSSLLPTF*KFT*KT*DD							
: : : : : : : :							
ED--DGAYIYEME-IQTKQGTETEFEISAKDGRRIKQEIDD							
5	140	150	160	170			

SEQ ID 8564 (GBS37) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 4; MW 22kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 10; MW 47kDa).

- 10 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 369

A DNA sequence (GBSx0400) was identified in *S.agalactiae* <SEQ ID 1203> which encodes the amino acid sequence <SEQ ID 1204>. Analysis of this protein sequence reveals the following:

15 Possible site: 53
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1499 (Affirmative) < succ>
 20 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9709> which encodes amino acid sequence <SEQ ID 9710> was also identified.

- 25 The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1205> which encodes the amino acid sequence <SEQ ID 1206>. Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 30 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2808 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 An alignment of the GAS and GBS proteins is shown below:

Identities = 128/297 (43%), Positives = 180/297 (60%), Gaps = 9/297 (3%)

40 Query: 54 IDDIKVGSPIFKYFWT-SLSLQAPLKALEFVLEQAKMPTELSGELSETQYLVAQFSDELA 112
 I D ++GSP F W Q+ + L F+L+ +MP ELSG+L ETQ L+ +F L
 Sbjct: 46 IIDNRLGSPTFWVIWPIEKENQSAKQLLTFLDLVEMPFELSGQLHETQTLLTRFHPSLL 105

Query: 113 PHDDFWIALSQVITYDSFPFGNSLAEDTVLNKRLHQFRYLISSQQAQYVRRYFKDVGMTDRD 172
 P FW L+ ++ +FPG +L++ L ++LHQFRY+ISSQQAQ +R ++K + MTD
 45 Sbjct: 106 PDHMFWKELASLVQAFPGKTLSQAGELEKRLHQFRYVISSQQAQSIERNHYKMIEMTDAQ 165

Query: 173 ALVNYL----SCL-REPDSIAYYESARLHNKRRRNGEIFGFPPDDEPVINSKLLISFHT 226
 AL +L CL R+ +SARLHNK R FP E N K+L+ FHT
 Sbjct: 166 ALALFLRSKKGPCLWRQAPDYTLMDSARLHNKLRFEDNKVIFPSQEVSYNIKVLLWFHT 225

50 Query: 227 FIIDDKGNFLNEIDADEVITRNGIINGASFNYAFKNNTTRHKELDVDPVK-LDPKFRNDMTR 285
 F +D G FLNE+DAEV+T GI+NGASFNY + RH +LDVDP+ DP+FR D +
 Sbjct: 226 FTLDSTGFFLNEVAEVTEKGIVNGASFNYG-TDGPRHWLDVDPISHHDQFRRDTLK 284

-470-

Query: 286 GYRSPNLSRRKWFFFKEEDYDCSYFNKKGYYAFGRRSAKQSVDKQVKYLKKAVQKMR 342
 G+RSP R+WF +++D+ SYFN KG +A+ +S+ V K K K+ + ++
 Sbjct: 285 GFRSPKRVFRQWFRAQKDDMFSYFNAKGLFAYHNKSSARVKKSAKQFKRQIHPPIK 341

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 370

- A DNA sequence (GBSx0401) was identified in *S.agalactiae* <SEQ ID 1207> which encodes the amino acid sequence <SEQ ID 1208>. This protein is predicted to be similar to two-component response regulator 10 [YcbM] (ompr-likeprotein). Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
15 bacterial cytoplasm --- Certainty=0.3129(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
20 >GP:CAA55264 GB:X78502 gtcR [Brevibacillus brevis]
  Identities = 99/228 (43%), Positives = 149/228 (64%), Gaps = 3/228 (1%)

  Query: 2 RTVLVVQGDDETIELLRSYLEGALYKVVMASDGEEAFSLFQQHQIDLAIIDITLPKIDGY 61
          +T+L+ + E IELL+ +LE Y+++ A DGE+A++ +QH +DLAIDI +P +DG+
 25 Sbjct: 3 KTILIAADDEPEIIIELLKLFLERESYRIIEAYDGEQAWNYIRQHPVVDLAIIDIMMPALDGF 62

  Query: 62 ELTRLIRODSQIPIIMLAAKTTDMDRILGLNIAGADDFITKPFNSLEVRLARINSQLRRYYE 121
          +L + + + ++P+I+L+AK D D+ILGL +GADDFT+KPFN LE +ARI +QLRR +E
  Sbjct: 63 QLIKRLTNEYKLPVIILSAKNRDSKDKILGLGLGADDFTSKPFNPNEAVARIQAQLRRAFE 122
30

  Query: 122 FNLSLAKP--KNQFKIGELELDEEHVELTKNGKHIKLTATEFKILHILMS-SPGRIYTKT 178
          FN + Q +G L L + + + +T E+++L+ M S I+TK
  Sbjct: 123 FNEPEEKAIYSTQSTTVGRILLLHTACVVYRGDETYSVTPLEYRLLNTFMQCSRTSIFTKQ 182

  Query: 179 QLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGYK 226
          QL+E+ D+ TIMV IS +RDKIED + P YIKT+RG+GYK
  Sbjct: 183 QLFEQAWSETYWEDDNTIMVQISRDLKDIEDQPRQPVVIKTVRGLGYK 230
```

There is also homology to SEQ ID 1182:

```
40   Identities = 87/230 (37%), Positives = 144/230 (61%), Gaps = 5/230 (2%)

  Query: 1 MRTVLVVQGDDETIELLRSYLEGALYKVVMASDGEEAFSLFQQHQIDLAIIDITLPKIDG 60
          M+ +L+V + +++++ L Y +V A DG EA ++F++ + DL I+D+ LP++DG
 45  Sbjct: 1 MKKILIVDDEKPISDIKFKNLTKEGYDIVTAFDGREATIFEEKPDLIILDLMLPELDG 60

  Query: 61 YELTRLIRODSQIPIIMLAAKTTDMDRILGLNIAGADDFITKPFNSLEVRLARINSQLRRYY 120
          E+ + IR+ S +PIIML+AK ++ D+++GL IGADD++TKPF++ E+LAR+ + LRR
  Sbjct: 61 LEVAKERIKTSHVPIIMLSADSEFDKVGLEIGADDYVTKPFNSRELLARVKAHLRRTE 120

  Query: 121 EFNSLAKPKN----QFIKIGELELDEEHVELTKNGKHIKLTATEFKILHILMSSPGRIY 175
          + +N Q + IG L++ + K+G+ ++LT EF++LH L + G++
  Sbjct: 121 TIETAVAEENASSGTQELTIGNLQILPDAFVAKKHGQEVELTHREFELLHHLANHMGQVM 180

  Query: 176 TKTQLYEKINGRYLEGDETTIMVHISNIRDKIEDDSKYPKYIKTLRGVGY 225
          T+ L E + G GD T+ V + +R+KIED P+YI T RGVGY
 55  Sbjct: 181 TREHLLIEIVWGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYILTRRGVGY 230
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 371

A DNA sequence (GBSx0402) was identified in *S.agalactiae* <SEQ ID 1209> which encodes the amino acid sequence <SEQ ID 1210>. This protein is predicted to be threonyl-tRNA synthetase 1 (thrS). Analysis of this protein sequence reveals the following:

5 Possible site: 32
 >>> Seems to have no N-terminal signal sequence

10 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.2353 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP: BAB06860 GB: AP001517 threonyl-tRNA synthetase 1 [Bacillus halodurans]
 15 Identities = 413/638 (64%), Positives = 506/638 (78%), Gaps = 7/638 (1%)

Query: 1 MIKITFPDGAIREFESGITTEIAQSISSNLAKKALAGKFNGQLIDTTRAIEEDGSIEIV 60
 MI ITFPDGA++EF G TT EIA SIS L KKALAG +G L+D IE+DG+I IV
 20 Sbjct: 4 MINITFPDGAKEFFPKGTTAAEIAGSISPGLKKKALAGMLDGTLDDLNTPIEQDGTITIV 63

Query: 61 TPDHEDALGVLRHSAAHLFAQAAKRLFPD--LCLGVGPQAIQDGFYDTDNKSQQISNDL 118
 TP+ ++AL VLRHS AH+ AQA KRLF D + LGVGP I+ GFYYD D ++ +DL
 Sbjct: 64 TPESDEALEVLRHSTAHVMAQALKRLFKDRNVKLGVGPVIEGGFYDVDMDES-LTPEDL 122

25 Query: 119 PRIEEEMKKIVKENHPCIREEISKEEAELELFKD--DPYKVELISEHAEDG-LTVYRQGEF 175
 P+IE+EMKKI+ EN P R +S+EEAL +++ DPYK+ELI++ ED +T+Y QGEF
 Sbjct: 123 PKIEKEMKKIIGENLPIERVVVSREEALARYEEVGDPYKIELINDLPEDETITIYEQGEF 182

30 Query: 176 VDLCRGPVHPSTGRIQVFHLLNVAGAYWRGNSDNAMMQRVYGTAWFDKDKDLKAYLKRREE 235
 DLCRG HVPSTG+++ F LLN+AGAYWRG+S N M+QR+YGT+A F K DL +L+ EE
 Sbjct: 183 FDLCRGVHVPSTGKILKEFKLNLLAGAYWRGDSNSKMLQRIYGTAFFKKADLDEHLRLLEE 242

35 Query: 236 AKERDHRKLGKELDLFMVNPEVGQGLPFWLPNGATIRRELERYIVDKEIASGYQHVTTP 295
 AKERDHRKLGKEL +F ++ +VGQGLP WLP GATIRR +ERYIVDKE GYQHVTTP
 Sbjct: 243 AKERDHRKLGKELGIFALSQKVQGQLPLWLPKGATIRRIIERYIVDKEEKLYQHVTTP 302

40 Query: 296 MASVEFYKTSGHWDHYREDMFPTMDMGDGEFFVLRPMNCPHIEVYKHHVHSYRELPARI 355
 +AS E YKTSQHWDHY++DMFPTM+M + EE VLRPMNCPH+ VYK + SYR+LP+RI
 Sbjct: 303 LASSELYKTSGHWDHYKDDMFPTMEM-ENEELVLRPMNCPHMMVYKTEMRSYRQLPLRI 361

45 Query: 356 AELGMMHRYEKSGALTGLQRVREMTLNDAHIFVTPEQIKDEFKALNLIAEIYEDFNLT 415
 AELG+MHRYE SGA++GLQRVRV MTLNDAHIF P+QIKDEF++ + LI +YEDF L +
 Sbjct: 362 AELGLMHRYEMSGAVSGLQRVRGMLNDAHIFCRPDQIKDEFVRRVRLIQAVYEDFGKLN 421

50 Query: 416 YRFRLSYRDPEDKHKYYDNDEMWNENAQAMLKEAMDDFGLDYFEAEGEAAFYGPKLDIQUK 475
 Y FRLSYRDPEDK KY+D+D MW AQ MLKEAMD+ L+YFEAEGEAAFYGPKLD+QV+
 Sbjct: 422 YSFRLSYRDPEDEKYYFDDDNMWNAQGMLKEAMDELELEYFEAEGEAAFYGPKLDVQVR 481

55 Query: 476 TALGNEETLSTIQLDFLLPERFDLKYGADGEEHPRIMIHRGGISTMERFTAILIETYKG 535
 TALG +ETLST+QLDFLLPERFDL Y+G DG+ HRP+++HRG +STMERF A L+E YKG
 Sbjct: 482 TALGKDETGSTVQLDFLLPERFDLTYVGEDGQPHRPVVHGVSTMERFVAFLLEEYKG 541

60 Query: 536 AFPTWLAPQQVSVIPISNEAHIDYAVEVARVLKDRGIRAEVDDRNEKMQYKIRAAQTQKI 595
 AFPTWLAP QV VIP+S EAH++YA V L+ GIR E+D+R+EK+ YKIR AQ QKI
 Sbjct: 542 AFPTWLAPQVQVIPVSPEAHLEYAKNVQETLQQAGIRVEIDERDEKIGYKIREAQMQKI 601

Query: 596 PYQLIVGDKEEMEEKAVNVRRYGSKATETKSIEEFVESI 633
 PY L++GDKE+E VNVR+YG K + + ++EFV +
 Sbjct: 602 PYMLVLDKEVEANGVNRKYGEKDSSMGLDEFVRHV 639

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1211> which encodes the amino acid sequence <SEQ ID 1212>. Analysis of this protein sequence reveals the following:

-472-

Possible site: 32
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2566 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 564/644 (87%), Positives = 608/644 (93%)

Query: 1 MIKITFPDGAIREFESGITTFEIAQSISNSLAKKALAGKFNGQLIDTTRAIEEDGSIEIV 60
 MIKITFPDGA+REFESG+TTF+IA+SIS SLAKKALAGKFN QLIDTTRAIEEDGSIEIV

15 Sbjct: 1 MIKITFPDGAIREFESGITTFDIAESISKSLAKKALAGKFNDQLIDTTRAIEEDGSIEIV 60
 Query: 61 TPDHEDALGVLRHSA AHLFAQAQAKRLFPDLCLGVGP AIQDGFFYD TDNKSGQI SNDLPR 120
 TPDH+DA VLRHSA AHLFAQAQAKRLFP+L LGVGPAI +GFYYDTDN GQISN+DLPR
 Sbjct: 61 TPDHKDAYEVLRHSA AHLFAQAQAKRLFPNLH LGVGPAIAEGF YYD TDNAEGQI SNEDLPR 120

20 Query: 121 IEEEMKKV KENHPCIREE EISKEE ALELFK DDPYKVEL ISEHA EDGL TVYR QGE FVD LCR 180
 IE EM+KIV EN+PCIREE++KEEAEALELFK DDPYKVELI+EHA GLTVYR QGE FVD LCR
 Sbjct: 121 IEAEMQ KIVT ENYPCIREEV TKEE ALELFK DDPYKVEL INEHAGA GLTVYR QGE FVD LCR 180

25 Query: 181 GPHVPSTGRI QVFH LLNVAGAYW RGN SDNAMM QRVY GTAWFD KDLKAY LKR REEA KERD 240
 GPHVPSTGRI QVFH LLNVAGAYW RGN SDNAMM QRVY GTAWFD KDLKAY L R EEA KERD
 Sbjct: 181 GPHVPSTGRI QVFH LLNVAGAYW RGN SDNAMM QRVY GTAWFD KDLKAY LTR REEA KERD 240

30 Query: 241 HRKL GKE LD LFMVN P EVGQ GLP FWL PNGAT IRREL ERYI VDKEIAS GYQH VYTP PMAS VE 300
 HRKL GKE LD LFM++ EVGQ GLP FWL +GATIRR LERYI DKE+ ASGYQH VYTPP+AS VE
 Sbjct: 241 HRKL GKE LD LFMISQ EVGQ GLP FWL PDGAT IRRT LERYI TDKE LAS GYQH VYTP PLAS VE 300

35 Query: 301 FYKT SGHWDHY REDMF PTMD MG DGE EFV LRP MNCP HIEV YKHHVHSY RELPIRIAELGM 360
 YKT SGHWDHY +EDMF P MDMG DGE EFV LRP MNCP HIEV YKHHVHSY RELPIRIAELGM
 Sbjct: 301 LYKT SGHWDHY QEDMF PVMD MG DGE EFV LRP MNCP HIEV YKHHVHSY RELPIRIAELGM 360

40 Query: 361 MHRYEKSGALTGLQRVREMTLND AHIFV TPEQIKDEFLK ALNLIAE IYED FNLT DYRF RL 420
 MHRYEKSGAL+GLQRVREMTLND HIFV TPEQI++EF +AL LI ++Y DFNL TDYRF RL
 Sbjct: 361 MHRYEKSGALSGLQRVREMTLNDGHIFV TPEQI QEEF QRAL QLI IDVYAD FNLT DYRF RL 420

45 Query: 421 SYRD PEDK HKYY DNDEM WENA QAML KEAM DD FG LDY FEA EGEA AFY GP KLDI QVK TAL GN 480
 SYRD P D HKYY DNDEM WENA QML A+A+ G+DY FEA EGEA AFY GP KLDI QVK TAL GN
 Sbjct: 421 SYRD PN DTHK YY DNDEM WENA QSM LKA AL DEM GVD Y FEA EGEA AFY GP KLDI QVK TAL GN 480

50 Query: 481 EETL STI QLD FLL PERFD LKYI GADGEE HRP IMH RG GIST MERFT AIL IETY KGAF PTW 540
 EETL STI QLD FLL PERFD LKYI GADGEE HRP +MIH RG IST MERFT AIL IETY KGAF PTW
 Sbjct: 481 EETL STI QLD FLL PERFD LKYI GADGEE HRP VMIH RG VIST MERFT AIL IETY KGAF PTW 540

55 Query: 541 LAPQQ SVIP ISNEAH IDY AWE VARVL KDRG IRAE VDDR NE KM QYK I RA QT QK I PY QL I 600
 LAP QV+VIP ISNEAH IDY AWE VA+ L+DRG+RA+VDDR NE KM QYK I RA+QT K I PY QL I
 Sbjct: 541 LAPHQV T VIPI SNEAH IDY AWE VAKTL DRGV RA DVDDR NE KM QYK I RA SQT SK I PY QL I 600

Query: 601 VGD KEM EKAVN VRRY GS KATE KSIE EFV ESI LADI ARKS RPD 644
 VGD KEM EK+VNR RY GS K T T+S+EEF VE+I LADI ARKS RPD
 Sbjct: 601 VGD KEM EDKS VNV RY GS KTT HTES VEEF VEN I LADI ARKS RPD 644

55 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 372

A DNA sequence (GBSx0403) was identified in *S.agalactiae* <SEQ ID 1213> which encodes the amino acid sequence <SEQ ID 1214>. Analysis of this protein sequence reveals the following:

Possible site: 16
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.1985 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA72250 GB:Y11463 ORF5 [Streptococcus pneumoniae]
 Identities = 189/290 (65%), Positives = 234/290 (80%)

10 Query: 1 MRIGLFTDTYFPQVSGVSTSIRTLKEGLEKEGHEVYIFTTDRNVKRKFEDPTIIRLPSVP 60
 MRIGLFTDTYFPQVSGV+TSIRTLK LEK+GH V+IFTTTD++V R+ED IIR+PSVP
 Sbjct: 1 MRIGLFTDTYFPQVSGVATSIRTLKTELEKQGHAVFIFTTDRDVNRYEDWQIIRIPSVP 60

15 Query: 61 FISFTDRVVYRGLISAYRIAKDYELDIHTQTEFSLGLLGKLVAKALRIPVVHTYHTQY 120
 F +F DRR YRG A IAK Y+LDIIHTQTEFSLGLLG +A+ L+IPV+HTYHTQY
 Sbjct: 61 FFAFKDERRFAYRGFSKALEIAKQYQLDIIHTQTEFSLGLLGIWIARELKIPVIHTYHTQY 120

20 Query: 121 EDYVGYYIAKGKLIKPSMVKYIMRTYLSLDGVICPSRIVLNLLDGYGVKIPKQVIPTGIP 180
 EDYV YIAKG LI+PSMVKY++R +L D+DGVICPS IV +LL Y VK+ K+VIPTGI
 Sbjct: 121 EDYVHYIAKGMLIRPSMVVKYLVRGFLHDVDGVICPSEIVRDLLSDYKVKEKRVIPGTIE 180

25 Query: 181 VENYRREDISEETIKNLRTELGLADNDTMLLSLSRVSFEKNIQAILMHL SAVVDENPHVK 240
 + + R +I +E +K LR++LG+ D + LLSSR+S+EKNIQA+L+ + V+ E VK
 Sbjct: 181 LAKFERPEIKQENLKERSKLGIQDGEKTLLSLSRISYEKNIQAVLVAFAADVKEEDKV 240

30 Query: 241 LVIVGDGPYLSDLKELVHSLELENSVIFTGMVEHSQVAIYYYKACDFFISA 290
 LV+ GDGPYL+DLKE +LE++SVIFTGM+ S+ A+YYKA DFFFISA
 Sbjct: 241 LVVAGDGPYLNDLKEQANLEIQDSVIFTGMIAPSETALYYKAADFFFISA 290

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1215> which encodes the amino acid sequence <SEQ ID 1216>. Analysis of this protein sequence reveals the following:

Possible site: 17

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.1074 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 309/444 (69%), Positives = 370/444 (82%)

45 Query: 1 MRIGLFTDTYFPQVSGVSTSIRTLKEGLEKEGHEVYIFTTDRNVKRKFEDPTIIRLPSVP 60
 MRIGLFTDTYFPQVSGV+TSIRTLK LEKEGHEVYIFTTDR+VKRKFEDPTIIRLPSVP
 Sbjct: 1 MRIGLFTDTYFPQVSGVATSIRTLKEELEKEGHEVYIFTTDRDVKRKFEDPTIIRLPSVP 60

50 Query: 61 FISFTDRVVYRGLISAYRIAKDYELDIHTQTEFSLGLLGKLVAKALRIPVVHTYHTQY 120
 F+SFTDRVVYRGLI+Y+I AK Y LDIIHTQTEFSLGLLGK++ KALRIPVVHTYHTQY
 Sbjct: 61 FVSFTDRVVYRGLISSYKIAKHYNLDIIHTQTEFSLGLLGKMIGKALRIPVVHTYHTQY 120

55 Query: 121 EDYVGYYIAKGKLIKPSMVKYIMRTYLSLDGVICPSRIVLNLLDGYGVKIPKQVIPTGIP 180
 EDYV YIA GK+I+PSMV ++R YL DLDGVICPSRIVLNLL+GY V IPK+VIPTGIP
 Sbjct: 121 EDYVSYIANGKIIIRPSMVKPLLRGYLKDLDGVICPSRIVLNLLLEGYEVТИPKRVIPTGIP 180

Query: 181 VENYRREDISEETIKNLRTELGLADNDTMLLSLSRVSFEKNIQAILMHL SAVVDENPHVK 240
 +E Y R+DI+ E + NL+ ELG+A ++TMLLSLSR+S+EKNIQA+ + A++ EN +K
 Sbjct: 181 LEKYIIRDITAEVTNLKAELGIAGDETMLLSLSRISYEKNIQAIINQMPAILAENAKIK 240

60 Query: 241 LVIVGDGPYLSDLKELVHSLELENSVIFTGMVEHSQVAIYYYKACDFFISASTSETQGLTY 300
 L+IVG+GPYL DLK L LE++ V FTGMV H +VA+YYKA CDFFISASTSETQGLTY
 Sbjct: 241 LIIVGNGPYLQDLKHLAMQLEVDKHVTFTGMVPHDKVALYYKA CDFFISASTSETQGLTY 300

Query: 301 IESLASGRPIIAQSNPYLDVVISDKMFGTLYKKESDLADAILDAIAETPKMTQEAYEQKL 360

-474-

IESLASG PIIA NPYLDDV++DKMFGTLY E+DL DAI+DAI +TP M + +K
 Sbjct: 301 IESLASGTPPIIAHGNPYLDDVVTDKMFGLYAAETDLTDAIIDAILKTPVMDKRLLAKKR 360

Query: 361 YEISAENFSKSVYAFYLDLISQKASVKEKVSLSLTIGNKDSHSTLRFVRKAVYLPKKVFTF 420
 YEISA++F KS+Y FYLD LI++ + +K+SL + + S+L+ V+ A++LPK+
 Sbjct: 361 YEISAQHFGKSIYTTFYLDTLIARNSKEAQKLSLYLNHSGKSSSLKLVQGAIHLPKRAAKV 420

Query: 421 TGRASKKVVVKAPKRISSIRDFLD 444
 T S KVVKAP + + +I+DFLD
 Sbjct: 421 TAITSVKVVKAPIKLVHAIKDFLD 444

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 373

A DNA sequence (GBSx0404) was identified in *S.agalactiae* <SEQ ID 1217> which encodes the amino acid sequence <SEQ ID 1218>. This protein is predicted to be lipopolysaccharide biosynthesis protein-related protein. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----
 bacterial cytoplasm --- Certainty=0.4076 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:AAG19110 GB:AE005009 Vng0600c [Halobacterium sp. NRC-1]
 Identities = 117/350 (33%), Positives = 178/350 (50%), Gaps = 29/350 (8%)

Query: 1 MKVLLYLEEAEYLKKSGIGRAIKHQEKALQIAGIDYTTNPT----- 41
 M+ L YLEA E L+ G+ A Q AL+ ++ P
 Sbjct: 2 MRALNYLEAAEALR-GGMVTATNQRAALETTDVEVETPWRAGDPVRSIGSLAAGGSCF 60

Query: 42 DDFDLVHMNTYGIRSWLLMSKAKKTGKKVIMHGHSSTEEDFRNSFIGSNLVSPLFKWYLCR 101
 FD+ H N G S + A++T +++H H T EDF SF GS+ ++P + YL
 Sbjct: 61 TAFDVAHCNLVPGPSAVARHARRTDTPLVLH AHLTREDFAQSFRGSSTIAPALEPYLRW 120

Query: 102 FYQKADAIITPTDYSKQLIKAYGIKKPIFVLSNGIDLSRYQRSEKESAFRHYFHLKDD 161
 FY +AD ++ P++Y+K +++AY + PI LSNG+DL Q E + R F L D
 Sbjct: 121 FYSQADLVLCPSEYTKDVLRAYPVDAPIRQLSNGVDLESMQYESFRADTRARFDL--DG 178

Query: 162 KVVMGAGLYFMRKGIDQFVEVAKMPDIRFIWFGETNKWVIPRKVRQIVTKQHPSNVTF 221
 VV G F RKG+ F E+ AK D F WFG ++ + P+NVTF
 Sbjct: 179 TVVYAVGEVFERKGLTMFCEL-AKATDHEFAFWFGPYDEGPQAGAATRKWVADPPANVTFT 237

Query: 222 GYIKGDVYEGAMSASDAFFPSREETEGIVVLEALASHQHVVLRDIPVYHGWVTE-DSVE 280
 GY++ A A D + FP++ E +GI VLEA+A + VVLRDIPV+ + T+ +
 Sbjct: 238 GYMEDK--RAAFGAGDIYLFPAKVENQGIAVLEAMACGKPVVLRDIPVFRFFFDTGEDCL 295

Query: 281 LATDVDFVKEKLDKVLSGKSDKIKEGYH---VAEERSIERIAHELASVYQ 327
 + + + F + +D++ + + G + AES S++RI ELAS+Y+
 Sbjct: 296 MCSTFEAFRDAIDRLADDPELRTLGENARETAESHSLDRIGEELASIYE 345

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1219> which encodes the amino acid sequence <SEQ ID 1220>. Analysis of this protein sequence reveals the following:

Possible site: 61
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

-475-

bacterial cytoplasm --- Certainty=0.4088 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

5 An alignment of the GAS and GBS proteins is shown below:

Identities = 236/332 (71%), Positives = 276/332 (83%)

```

Query: 1 MKVLLYLEAEEYLKKSGIGRAIKHQEKALQIAGIDYTTNPTDDFLVHMNTYGIRSWLLM 60
      MKVLLYLEAEEYLKKSGIGRAIKHQ KAL + G +TTNP + +DLVH+NTYG++SWLLM
10   Sbjct: 1 MKVLLYLEAENYLRKSGIGRAIKHQAKALSIVGQHFTTNPRETYDLVHLNTYGLKSLLM 60

Query: 61 SKAKKTGKKVIMHGSTEEDFRNSFIGSNLVSPLFKWYLCRFYQKADAIITPTDYSKQLI 120
      KA+K GKKVIMHGSTEEDFRNSFI SNL+SP FK YLC FY KADAIITPT YSK LI
15   Sbjct: 61 IKAQKAGKKVIMHGSTEEDFRNSFIFSNLSPWFKKYLCHFYNKADAIITPTLYSKSLI 120

Query: 121 KAYGIKKPIFVLSNGIDLRSYQRSEKKESAFRHYFHLSKDDKVVMGAGLYFMRKGIDQFV 180
      ++YG+K PIF +SNGIDL +Y KKE+AFR YF + + +KVMGAGL+F+RKGID FV
      Sbjct: 121 ESYGVKSPIFAVSNGIDLEQYGADPKKEAAFRRYFDIKEGEKVVVMGAGLFFLRGIDDFV 180

20   Query: 181 EVAAKMPDIRFIWFGETNKWVIPRKVRQIVTKQHPSNVTFAAGYIKGDVYEGAMSASDAFF 240
      +VA MPD+RFIWFGETNKWVIP +VRQ+V HP N+ F GYIKGDVYEGAM+ +DAFF
      Sbjct: 181 KVAQAMPDVRFIWFGETNKWVIPAQRQMVNNGHPKNLIFPGYIKGDVYEGAMTGADAFF 240

25   Query: 241 FPSREETEGIVVLEALASHQHVVLRLDIPVYHGWVTEDSVELATDVDGFVEKLDKVLSGKS 300
      FPSREETEGIVVLEALAS QH+VLRDIPVY+GWV + S ELATD+ GF+E L KV SG S
      Sbjct: 241 FPSREETEGIVVLEALASRQHVLRLDIPVYYGWVDQSSAELATDIPGFIEALKVFSGAS 300

Query: 301 DKIKEGYHVAESRSIERIAHELASVYQKVMEL 332
      +K++ GY VA+SR +E + H L VY+KVMEL
30   Sbjct: 301 NKVEAGYKVAQSRRLETVGHALVDVYKKVME 332
  
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 374

35 A DNA sequence (GBSx0405) was identified in *S.agalactiae* <SEQ ID 1221> which encodes the amino acid sequence <SEQ ID 1222>. Analysis of this protein sequence reveals the following:

```

Possible site: 16
>>> Seems to have no N-terminal signal sequence

40 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.5487 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
  
```

45 The protein has homology with the following sequences in the GENPEPT database:

```

>GP: AAC35010 GB: AF055987 intracellular a-amylase [Streptococcus mutans]
  Identities = 308/483 (63%), Positives = 378/483 (77%)

50   Query: 1 MTNELIMQAWEWYLPSDLGNHWKKLEESISDLKLGISKIWLPPAFKGTSSDDVGYGVYDL 60
      MTNE +MQ FEWYLP+DG HW+ L E S LK +GISK+W+PPAFKGT S+DVGYGVYDL
      Sbjct: 1 MTNETMMQYFEWYLPNDGKHQHLAEDASHLKNIGISKVWMPPAFKGTGSNDVGYGVYDL 60

Query: 61 FDLGEFDQNGTIRTKYGRKEEYLKLICKSLKANGIKPFADIVLNHKANGDHKEKFQVIKVN 120
      +DLGEF+QNGT+RTKYG +E+YL + +LK I P +DIVLNHKANGD KE+FQV+KVN
55   Sbjct: 61 YDLGEFNQNGTVRTKYGREDYLNANKEQEIMPISDIVLNHKANGDAKERFQVVKVN 120

Query: 121 PENRQEALSEPYEIEGWTGFDFPGRQGEYNDFKWHWYHFTGLDYDAKNNETDIFMIVGDN 180
      P NRQE +SEPYEIEGWT F+FPGRQ Y+DFKWHWYHFTG+DYDA +NE I+MI+GDN
      Sbjct: 121 PSNRQEKEISYPYEIEGWTQFNFPGRQDNYSDFKWHWYHFTGVDYDALHNENGIYMLGDN 180

60   Query: 181 KGWADDLIDDENGNFYLMYNDIDFKHPEVIKNLQDWAKWFIETTGIEGFRLDIVKHID 240
  
```

-476-

KGWA + ID ENGN+DYLMY+DIDFKHPEV ++L+DW WF+ET+G+ GFRLDA+KHID
 Sbjct: 181 KGWASQENIDQENGNYDYLMYDDIDFKHPEVQEHLRDWVAWFLETSGVGGFRLDIAKHID 240

Query: 241 SYFIQTFINDIRTKIKPDLEVFGHEYWKSDQTSMKDYLEATQFQFSLVDVTLHMNFFDASH 300
 F+ FI IR +K DL VFGEYWK + DYL + QF L+DV LHM+ F+A
 Sbjct: 241 KTFMAQFIRYIREHLKADLYVFGEYWKDSHFIDTDYLHSVDLQFDLIDVMLHMSLFEAGQ 300

Query: 301 QNRDFDMRTIFDDSLVIDNPEYAVTFVENHDTQSGQALESRVEDWFKPLAYGLILLRQQG 360
 + DFD+ TI DDSL+ P++AVTFV+NHD+Q GQALES V +WFKPLAYGLILLRQ+G
 Sbjct: 301 KGSDFDLSTILDDSLMKSHPDFAVTFVDNHDSQRGQALESTVAEWFKPLAYGLILLRQEG 360

Query: 361 TPCLFYGDYYGIQGEFGQPSFKEVIDKMAELRQNYYFGKQVDYFTHSNCIGWTCLGDEEH 420
 PC+FYGDYYGI GEF Q SF+ V+DK+ +RQ +V+G + T NCIGWTCLGDEEH
 Sbjct: 361 IPCVFYGDYYGISGEFAQESFQTVLDKLLYIRQYHVYGSKKIILTMCNCIGWTCLGDEEH 420

Query: 421 NSCLAVVLTNGDQGWKHMEVGEIYAGKTFVDYLGNCQEVEVIGDDWGDFLIVESASISAW 480
 +AV+++NG+ K M +GE K FVDYL NC +EV++ D GWGDF V+ AS+SAW
 Sbjct: 421 PDGVAVIISNGEANCKRMNMGEFNRNKVFVDLNCTEEVILDDQGWGDFPVQEASLSAW 480

Query: 481 VPK 483
 V K
 Sbjct: 481 VNK 483

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1223> which encodes the amino acid sequence <SEQ ID 1224>. Analysis of this protein sequence reveals the following:

Possible site: 30

>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----

bacterial outside --- Certainty=0.3000(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAB00845 GB:M57692 alpha-cyclodextrin glycosyltransferase
 [Thermoanaerobacterium thermosulfurigenes]
 Identities = 356/710 (50%), Positives = 468/710 (65%), Gaps = 16/710 (2%)

Query: 7 KTYKLLTKSAVLLGLISFPLT--VSAADNASVTNKADFSTDITYQIVTDRFNDGNTSNNG 64
 KT+KL+ + L L+ F LT + AA + +V+N ++STD IYQIVTDRF DGNTSN
 Sbjct: 3 KTFKLILVLMLSITLV-FGLTAPIQAASDTAVSNVVNYSTDVIYQIVTDRFVDGNTSNNP 61

Query: 65 KTDVFDKN--DLKKYHGGDWQGIIAKIKDGYLTDGMISAIWISSPVENIDSIDPSN--G 119
 D++D LKKY GGDWQGII KI DGYLT MG++AIWIS PVENI ++ P + G
 Sbjct: 62 TGDLYDPTHTSLKKYFGGDWQGIIINKINDGYLTGMGVTAIWISQPVENIYAVLPDSTFGG 121

Query: 120 SAAYHGywAKDFFKTQNQHFGTEADFQQLVKVAHQHHIKVVIDFAPNHTSTAEGTTFKE 179
 S +YHGWA+DF +TN +FG+ DFQ L+ AH H+IKV+IDFAPNHTS A + T+ E
 Sbjct: 122 STSYHGyWARDFKRTNPYFGSFTDFQNLINTAHAHNIKVIIDFAPNHTSPASETDPTYAE 181

Query: 180 DGALYKNGKLVGKFSDDKDKIFNHESWTDFSTYENSIYHSMYGLADLNNINPKVDQYMKE 239
 +G LY NG L+G +++D + F+H TDFS+YE+ IY +++ LADLN N +D Y+K
 Sbjct: 182 NGRLYDNGTLLGGYTNDTNGYFHHYGGTDSSYEDGIYRNLFDLADLNQQNSTIDSYLNKS 241

Query: 240 AIDKWLDLGVDGIRVDAVKHMSQGWQKNWLISHIYEKHNVFVFGEWFSGHTDDYDMTTFA 299
 AI WLD+G+DGIR+DAVKHM GWQKN++ I VF FGEWF G + D + T FA
 Sbjct: 242 AIKVWLDMGIDGIRLDAVKHMPFGWQKNFMDSILSYRPVFTFGEWFLGTNEIDVNNTYFA 301

Query: 300 NNSGMGLLDFRFANAIROLYTGFSTFTMRDFYKVLENRDQVTNEVTDQVTFIDNHDMERF 359
 N SGM LLDFRF+ +RQ++ +T TM +++++ N + D VTFIDNHDM+RF
 Sbjct: 302 NESGMSSLDFRFQKVQRQVFRD-NTDTMYGLDSMIQSTASDYNFINDMVTFIDNHDMDRF 360

Query: 360 ATKVANNQTAVNQAYALLLTSRGVPNIYYGTEQYATGDKDPNNRGDMPSFNKESQAYKVI 419
 + V QA A LTSRGVP IYYGTEQY TG+ DP NR M SFN + AY VI

-477-

Subjct: 361 YN- GGSTRPVEQALAFTLTSRGVPAIYYGTEQYMTGNGDPYNRAMMTSFNTSTTAYNVI 418

Query: 420 SKLAPLRKQNQALAYGTTQRWIISDHVLVFERKFGNHNVALVAINRQDQTNGYTITNAKTL 479
KLAPLRK N A+AYGTT+QRWI++ V ++ERKFGN+VALVAINR+ + Y IT TAL

Sbj ct : 419 KKLAPLRSKNPAIAYGTTQQRWINNDVYIYERKFGNNVALVAIRNLSTSNTGLYTAL 478

Query: 480 PQNSYKDKLEGLLGGQELIVGADGTISSFELGAGQVAVWTYEGEDKTPQLGDVDASVGIA 539
P +Y D L GLL G + V +DG+++ F L AG+VA VW Y +P +G V ++ A

Subjct : 479 PAGTYTDVLGGLLNGNSISVASDGSVTPFTLSAGEVAVWQYVSSNSPLIGHVGPTMTKA 538

10

Query: 540 GNKITISGQFGNSKGQVTFGEISAEILSWSDTLITLKPTVPANYNISVTTADKQTSN 599
G ITI G+GFG + GQV FG + I+SW DT + +KVP+V YNIS+ T+ TSN

Subjct : 539 GQTITIDGRGFRTSGQVLFGSTAGTIVSWDDTEVKVKVPSVTPGKYNISLKTSSGATSN 598

15 Query: 600 SYQAFEVLDKQI PVRLLINDFKTVPGEQLYLMGDVFEMGANDAKNAVGPLFNNTQTIAK 659
+Y +LT QI VR ++N+ TV GE +YL G+V E+G D A+GP+FN Q + +

Subjct: 599 TYNNINITGNQICVRFVNNASTVYGENVYLTGNVAELGNWDTSKAIGPMFN--QVYQ 656

Query: 660 YPNWFFDTHLPINKEIAVKLVKKDSIGNVLWT--SPETYSIKTGHEAQTI 707

YP W++D +P I K +KK+ + W S TY++ + I

Sbjct: 657 YPTWYYDVSVPAGTTIQFKFIKKNG-NTITWEGGSNHTYTVPSSSTGTVI 708

An alignment of the GAS and GBS proteins is shown below:

Identities = 112/509 (22%), Positives = 193/509 (37%), Gaps = 103/509 (20%)

25 Query: 18 GHNWKLEESISD--LKKLGISKIWLPPAFKGTSDDV-----GYGVYDLFDLGEFD 67
 G W+ + I D L +GIS IW+ + S D GY D F +
 Shict: 79 GGGDWGCTTAKTKDGVLTDMGTSATIWSSPVENIDSTDRSNGSAAYHGIVWAKDEEFKTNQH- 137

30 Query: 68 QNGTIRTQYGRKEEYLKLIKSILKANGIKPFA DIVLNHKGANGDHKEFQVIVKVNPE NRQEAA 127
+G + ++ +L+K + IK D NH + + +
Sbjct: 138 -----FGTEADFFQQLVKVAHQHHIKVVVIDFPNHTSTAEKE----- 173

Query: 128 LSEPYEIEGWTGFDFPGRQGEYNDFKWHWYHFTGLDYDAKNNETDIFMIVGDNKGWADDD 187
 G F Y + K G D K+ + +++ W D
 Shift: 174 CTTTKEKFGALYKACK LUVKESPPDKDK LUVKESPPDKDK 210

Query: 188 LIDDE--NGNFDYLMYNDIFDKHPEVIKLNQLDWAKWFIEETTGIEGFRLDAVKHIDSYFIQ 245
++ + + N+I+ K + +K D KW G++G R+DAVKH+ + +
40 Sbjct: 211 TYENSTIYHSMYGLADLNNINPKVDQYMKEAID--KWL--DLGVGDGIRVDAVKHMSOGWQK 266

Query: 246 TFINDIRTKIKPDLEVFGEWKSQTSMDYLETAQFQFSLVDVTLHMNFFDASHQ-NRD 304
 +++ I K ++ VFGE W S T D + T F + L F+A Q
Shift: 267 NWLSHIVTE ..KUMVEVEGE-WECCHTD--DDYDMTTEANNSCMCTLDEPANATPOLYTC 321

45 Query: 305 FDMRTIFDDSLVIDNPEYA-----VTFVENHDTQSGQALESRVEDWFKPLAYGLLLR 357
F T+ D V++N + VTF++NHD + + + AY L LL
sbjct: 322 FSTFTMRDFYKVLENRDOVTNEVTDQVFIDNHDMERFATKVANNOTAVNO-AYAL-LLT 379

50 Query: 358 QQGTPCLFYGYGIQGE-----FGQPSFK-----EVIDKMAELR---QNYVFGKQVD 402
+G P ++YG G+ PSF +VI K+A LR Q +G
Sbjct: 380 SRGVPIIYGGTEQYATGDKDPNNGRDMPSPFNKESQAYKVISKLAPLRKQNQALAYGTTEQ 439

55 Query: 403 YFTHNSCNGWTCGLGDEEHNSCLAVVLNTGDQ--GWKHMEVGEIYAGKTFVDYLGN--EQ 458
+ + + + + + +A+V N DQ G+ ++ D L Q
Sbjct: 440 RWTSDHVL---VEERKFGNHVATVA1NRDNTNGYTITNAKTA1PONSYKDKELEGJLGG 495

Query: 459 EVVIGDDGW-GDFLVESASISIAWVPKIEE 486
E+++G DG F + + ++ W + E+
60

Based on this analysis, it was predicted that these proteins and their epitopes could be used for vaccines or diagnostics.

Example 375

A DNA sequence (GBSx0406) was identified in *S.agalactiae* <SEQ ID 1225> which encodes the amino acid sequence <SEQ ID 1226>. This protein is predicted to be catabolite control protein A. Analysis of this protein sequence reveals the following:

5 Possible site: 29
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 10 bacterial cytoplasm --- Certainty=0.2154 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9707> which encodes amino acid sequence <SEQ ID 9708> was also identified.

15 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAA88121 GB:AB028599 catabolite control protein A [Streptococcus bovis] (ver 3)
 Identities = 304/332 (91%), Positives = 320/332 (95%)

 20 Query: 1 MNTDDTTIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60
 MNTDDTTIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA
 Sbjct: 1 MNTDDTTIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60

 25 Query: 61 SKKTTTGVVVIPNIANSYFSILARGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120
 SKKTTTGVVVIPNIANSYFSILA+GIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ
 Sbjct: 61 SKKTTTGVVVIPNIANSYFSILAKGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120

 30 Query: 121 VDGIIFMGHHLTEKIRAEFSRSRTPIVLAGTVDEHQLPSVNIDYKAAADVVIDILAGNH 180
 VDGIIFMGHHLTEKIRAEFSRSRTP+VLAGTVDEHQLPSVNIDYKAA DV+DILA N+
 Sbjct: 121 VDGIIFMGHHLTEKIRAEFSRSRTPVLAGTVDEHQLPSVNIDYKAAADVADVVDILAKNN 180

 35 Query: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLKKNGLNFKEGLVFEANYRYAEGFALAQRVINAG 240
 KDIAFVSGPLIDDINGKVRLAGYKEGL+KN L+FKEGLVFEANY Y +G+ LAQRV+N+G
 Sbjct: 181 KDIAFVSGPLIDDINGKVRLAGYKEGLEKNNLSFKEGLVFEANYNYKDGYELAQRVMNSG 240

 40 Query: 241 ATAAYVAEDELAAGLLNGLFEAGKRPEDFEIITSNDSPPIAQYTRPNLTSISQPVYDLGA 300
 ATAAYVAEDELAAGLLNGLF AGK+VPEDFEI+TSNDSPPI YTRPNL+SISQPVYDLGA
 Sbjct: 241 ATAAYVAEDELAAGLLNGLFAAGKKVPEDFEIITSNDSPITSYTRPNLSSISQPVYDLGA 300

 45 Query: 301 VSMRMLTKIMHKEELEEKEIVLNHGIVKRGTT 332
 VSMRMLTKIM+KEELEEKEI+LNHG+ RGTT
 Sbjct: 301 VSMRMLTKIMNKEELEEKEIIINHGLKLRGTT 332

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1227> which encodes the amino acid sequence <SEQ ID 1228>. Analysis of this protein sequence reveals the following:

Possible site: 29
 >>> Seems to have no N-terminal signal sequence

 ----- Final Results -----
 50 bacterial cytoplasm --- Certainty=0.2154 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

55 Identities = 307/332 (92%), Positives = 320/332 (95%)

 Query: 1 MNTDDTTIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60
 MNTDD +TIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA
 Sbjct: 1 MNTDDPLTIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLA 60

Query: 61 SKKTTTGVVIPNIANSYFSILARGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120
SKKTTTGVVIPNIANSYFSILA+GIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ
Sbjct: 61 SKKTTTGVVIPNIANSYFSILAKGIDDIAAMYKYNIVLASSDEDDDKEVNVVNTLFAKQ 120

5

Query: 121 VDGIIIFMGHHHLTEKIRAEFSRSRTPIVLAGTVDLEHQLPSVNIDYKAAAVDVIDILAGNH 180
VDGIIIFMGHHHLTEKIRAEFSRSRTP+VLAGTVDL+HQLPSVNIDY+AA +V+DILA NH
Sbjct: 121 VDGIIIFMGHHHLTEKIRAEFSRSRTPVVLAGTVLDLHQLPSVNIDYRAAVSNVVNDILAENH 180

10

Query: 181 KDIASFVSGPLIDDINGKVRLAGYKEGLKKNGLNFKEGLVFEANYRYAEGFALAQRVINAG 240
K IAFVSGPLIDDINGKVRLAGYKEGLK N L+FKEGLVFEANY Y EGF LAQRVIN+G
Sbjct: 181 KCIAFVSGPLIDDINGKVRLAGYKEGLKHNLDFKEGLVFEANYSYKEGFELAQRVINSG 240

15

Query: 241 ATAAYVAEDELAAGLLNGLFEAGKRPEDFEIITSNDSPIAQYTRPNLTSISQPVYDLGA 300
ATAAYVAEDELAAGLLNGLFEAGKRPEDFEIITSNDSP+ QYTRPNL+SISQPVYDLGA
Sbjct: 241 ATAAYVAEDELAAGLLNGLFEAGKRPEDFEIITSNDSPVVQYTRPNLSSISQPVYDLGA 300

20

Query: 301 VSMRMLTKIMHKEELEEKEIVLNHGIVKRGT 332
VSMRMLTKIM+KEELEEKEI+LNHG I KRGT
Sbjct: 301 VSMRMLTKIMNKEELEEKEILLNHGIKKRGT 332

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 376

25 A DNA sequence (GBSx0407) was identified in *S.agalactiae* <SEQ ID 1229> which encodes the amino acid sequence <SEQ ID 1230>. This protein is predicted to be PepQ (pepQ-2). Analysis of this protein sequence reveals the following:

Possible site: 22
>>> Seems to have no N-terminal signal sequence
30 ----- Final Results -----
bacterial cytoplasm --- Certainty=0.1118 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

35 The protein has homology with the following sequences in the GENPEPT database:

>GP: AAC46293 GB: AF014460 PepQ [Streptococcus mutans]
Identities = 257/359 (71%), Positives = 304/359 (84%)

40 Query: 1 MSKLNIRHHLHSVQAELAVFSDPVTVNYLTGFFCDPHERQMFLFVYEDRDPILFVPALE 60
MSKL +I L E AV SDPV++NYLTGF+ DPHER MFLF++ D++ +LF+P L+
Sbjct: 1 MSKLAQIVQKLKKQGIEAAVLSDPVSINYLTGFYSDPHERLMFLFLFADQETLLFLPELD 60

45 Query: 61 VSRAKQSVPFPVFGYIDSENWPWKIASNLPSFSVSKVLAEDNLNVTKFQGLQTVDGHF 120
RAK + V GY+D ENP +KI + LP + SK+ EFDNLNVTKF+GL+T+F G F
Sbjct: 61 ALRAKSILDISVTGYLDFENPLEKIKTLLPKTNYSKIALEFDNLNVTKFKGLETFSGQF 120

50 Query: 121 ENLTPYIQNMRLIKSRDEIEKMLVAGEFADKAVQVGFDNISLNNTETDIIAQIEFEMKKQ 180
NLTP I MRLIKS DEI+K+L+AGE ADKAVQ+GFD+ISLN TETDIIAQIEFEMKK
Sbjct: 121 TNLTPLINRMRLIKSADEIQLKLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180

55 Query: 181 GINKMSFDTMVLITGNNAANPHGIPGTNKIENNALLLFDLGVE TLGYTSMDTRTVAVGKPD 240
G++KMSF+TMVLITG+NAANPHG+P ++KIENN LLLFDLGVE+ GY SDMTRTRVAVG+PD
Sbjct: 181 GVDKMSFETMVLITGSNAANPHGLPASHKTIENNHLFFDLGVESTGYVSDMTRTVAVGQPD 240

Query: 241 QFKKDIYHLCLEAHQAAIDFIKPGVLASEVDAARNVIEKAGYQGYFNHRLGHGLGMDVH 300
QFKKDIY++CLEA A+DFIKPGV A++VDAAR+VIEKAGY Y FNHRLGHG+GM +H
Sbjct: 241 QFKKDIYNICLAEQLTALDFIKPGVSAAQVDAARSVIEKAGYGDYFNHRLGHGIGMGLH 300

60 Query: 301 EFP SIMAGNDMEI QEGMCF SVE PGIYIP DKVGVRI EDCG YVTKTGFEVFTKTPKELLYF 359
EFP SIMAGNDM +EGMCF SVE PGIYIP+KVGVRIEDCG+VTK GFEVFT+TPKELLYF

-480-

Sbjct: 301 EFPSIMAGNDMLLEEGMCFSEPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1231> which encodes the amino acid sequence <SEQ ID 1232>. Analysis of this protein sequence reveals the following:

5 Possible site: 58
 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -0.90 Transmembrane 42 - 58 (42 - 59)
 10 ----- Final Results -----
 bacterial membrane --- Certainty=0.1362 (Affirmative) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

15 >GP:AAC46293 GB:AF014460 PepQ [Streptococcus mutans]
 Identities = 264/359 (73%), Positives = 304/359 (84%)
 Query: 1 MTKLDQIRLYLDQKGAEALAI FSDPVTINYLTGFFCDPHERQLFLFVYHDLAPVLFVPALE 60
 M+KL QI L ++G E A+ SDPV+INYL TGF+ D PHER +FLF++ D +LF+P L+
 20 Sbjct: 1 MSKLAQIVQKLKKQGIEAAVLSDPVSINYLTGFYSDPHERLMFLFLFADQETLLFLPEILD 60
 Query: 61 VARASQAISFPVFGYVDSEN PWEKIKAVLPNTAAKTIYAEFDHNLNVNKFHGLQTIFSGQF 120
 RA + V GY+D ENP EKIK +LP T I EFD+LNV KF GL+TIFSGQF
 Sbjct: 61 ALRAKSILDISVTGYLD FENPLEKIKTLLPKTNYSKIALEFDNLNVTKFKGLETI FSGQF 120
 25 Query: 121 NNLT PYVQGMRLVKSAD EINKMMIAGQFADKAVQVGF DNISLDATETDVIAQIEFEMKKQ 180
 NLTP + MRL+KSADEI K++JAG+ ADKAVQ+GFD+ISL+ATETD+IAQIEFEMKK
 Sbjct: 121 TNLTPLINRMRLIKSAD EIQKLLIAGELADKAVQIGFDSISLNATETDIIAQIEFEMKKL 180
 30 Query: 181 GIHKMSFDTMVL TGNNNAANPHGIPGTNNIENNALLFDLG VETLGYTSMDTRTVAVGQPD 240
 G+ KMSF+TMVL TG+NAANPHG+P ++ IENN LLLFDLGVE+ GY SDMTRTVAVGQPD
 Sbjct: 181 GVDKMSFETMVL GSNAANPHG LPA SHKIEN NHLLFDLG VESTGYVSDMTRTVAVGQPD 240
 35 Query: 241 QFKID IYNLCLEAQLAAIDFIKPGV TAAQVDAARQVIEKAGYGEYFNHRLGHGIGMDVH 300
 QFK DIYN+CLEAQL A+DFIKPGV+AAQVDAAR VIEKAGY G+YFNHRLGHGIGM +H
 Sbjct: 241 QFKKDIYNICL EAQLTALDFIKPGV SAAQVDAAR SVIEKAGY GDYFNHRLGHGIGMGLH 300
 40 Query: 301 EFPSIMAGNDLV LEEGMCFSEPGIYI PPGKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359
 EFPSIMAGND++ LEEGMCFSEPGIYI PPGKVGVRIEDCGHVTKNGFEVFT QTPKELLYF
 Sbjct: 301 EFPSIMAGNDMLLEEGMCFSEPGIYIPEKVGVRIEDCGHVTKNGFEVFTQTPKELLYF 359

An alignment of the GAS and GBS proteins is shown below:

Identities = 288/361 (79%), Positives = 325/361 (89%)

45 Query: 1 MSKLNRIRHHHLHSVQAE LAVFSDPVTVN LTGFFCDPHERQMFLFVYEDRDPILFVPALE 60
 M+KL++IR +L AELA+FSDPVT+NYLTGFFCDPHERQ+FLFVY D P+LFVPALE
 Sbjct: 1 MTKLDQIRLYLDQKGAEALAI FSDPVTINYLTGFFCDPHERQLFLFVYHDLAPVLFVPALE 60
 50 Query: 61 VSRAKQSVFPVFGYIDSEN PWEKIKAVLPNTAAKTIYAEFDHNLNVNKFHGLQTIFSGQF 120
 V+RA Q++ FPVFGY+DSEN PWP+KI + LP+ + AEFD+LNV KF GLQT+F G F
 Sbjct: 61 VARASQAISFPVFGYVDSEN PWEKIKAVLPNTAAKTIYAEFDHNLNVNKFHGLQTIFSGQF 120
 Query: 121 ENLTPYI QNMR LIKSR DEIEKMLVAGEFADKAVQVGF DNISLNNTETDIIAQIEFEMKKQ 180
 NLTPY+Q MRL+KS DEI KM++AG+FADKAVQVGF DNISL+ TETD+IAQIEFEMKKQ
 55 Sbjct: 121 NNLT PYVQGMRLVKSAD EINKMMIAGQFADKAVQVGF DNISLDATETDVIAQIEFEMKKQ 180
 Query: 181 GINKMSFDTMVL TGNNNAANPHGIPGTNNIENNALLFDLG VETLGYTSMDTRTVAVGKPD 240
 GI+KMSFDTMVL TGNNNAANPHGIPGTN IENN ALLLFDLG VETLGYTSMDTRTVAVG+PD
 Sbjct: 181 GIHKMSFDTMVL TGNNNAANPHGIPGTNNIENNALLFDLG VETLGYTSMDTRTVAVGQPD 240
 60 Query: 241 QFKKDIYHLCLEAQLAAIDFIKPGV LASEVDAAR NVIEKAGY GQYFNHRLGHGIGMDVH 300
 QFK DIY+LCLEA AAIDFIKPGV A++VDAAR VIEKAGY G+YFNHRLGHG+GMDVH
 Sbjct: 241 QFKID IYNLCLEAQLAAIDFIKPGV TAAQVDAARQVIEKAGYGEYFNHRLGHGIGMDVH 300

-481-

Query: 301 EFPSIMAGNDMEIQEGMCF SVEPGIYIPDKGVRIEDCGYVTKTGFEVFTKTPKELLYFEG 361
 EFPSIMAGND+ ++EGMCF SVEPGIYIP KVGVRIEDCG+VTK GFEVFT TPKELLYFEG
 Sbjct: 301 EFPSIMAGNDLVLEEGMCF SVEPGIYIPGKGVRIEDCGHVTKNGFEVFTHTPKELLYFEG 361

- 5 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 377

A DNA sequence (GBSx0408) was identified in *S.agalactiae* <SEQ ID 1233> which encodes the amino acid sequence <SEQ ID 1234>. Analysis of this protein sequence reveals the following:

10 Possible site: 14
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 15 bacterial cytoplasm --- Certainty=0.3629 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

- 20 Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 378

A DNA sequence (GBSx0409) was identified in *S.agalactiae* <SEQ ID 1235> which encodes the amino acid sequence <SEQ ID 1236>. This protein is predicted to be beta-hexosaminidase A precursor. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----
 30 bacterial cytoplasm --- Certainty=0.3279 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

35 >GP:CAB11942 GB:Z99104 alternate gene name: yzbA-similar to
 beta-hexosaminidase [Bacillus subtilis]
 Identities = 151/602 (25%), Positives = 268/602 (44%), Gaps = 69/602 (11%)

40 Query: 26 INEMTLDEKIGQLF-----FNMGASRSEEEYLTDVLDRYHIAAVRYNRGSSEIYDQNL- 78
 +N M+LDEK+GQ+ + S + LT + D +Y G ++ +N+
 Sbjct: 39 VNRMMSLDEKLGQMLMPDFRNWQKEGESSPQALTkmndevaslvkkyqfggii-LFAENVK 97

45 Query: 79 -----ILQTKSKLPLMIIAANTEAGGDGAVIDGTVKGDEIKVAATNDPKYAYEMG 127
 + K+P++++ + E G + +GT + + A AY+ G
 Sbjct: 98 TTKQTVQLTDDYQKASPKIPLMLSIDQEGGIVTRLGETNFPGNMALGAARSRINAYQTG 157

50 Query: 128 RIAGMEASAVGCNASFSPIVDLTRNWRNPPIIASRNWGANVDQIISLSKEYMKGIMQYNIV 187
 I G E SA+G N FSP+VD+ N NP+I R++ +N + L MKG+ + +I
 Sbjct: 158 SIIGKELSALGINTDFSPVVDDINNNPDNPVIGVRSFSSNRELTSRLGLYTMKGLQRQDIA 217

Query: 188 PFAKHFPGDGIDERDHLSFASNPMSKEEWMSTFGRIYGEIADAGLPGVMAGHIHLPNVE 247
 KHFPG G + D H +E + + DAG VM H+ P +
 Sbjct: 218 SALKHFPGHGDTDVDSHYGLPLVSHGQERLREVELYPFQKAIDAGADMVMTAHVQFPAPD 277

-482-

Query: 248 KEMHPER--DLDDMLPASLNKTLLDELLRGELEYNGAIVTDASHVGMTASMARRDLLPT 305
 + + D ++PA+L+K ++ LLR E+G+NG IVTDA +M + + + +

Sbjct: 278 DTTYKS KLDGS DILVPATLSKKVMTG LRLQEMGF NGV I VTD ALNMKA IADHFG QEEAVVM 337

5 Query: 306 AIEAGCDLFLF---FNPDDED-----IQWMKEGYEK GILTEERLHDALRRTLGLKAKLG 356
 A++AG D+L E+ IQ +KE + G + E++++++ R + LK K G
 Sbjct: 338 AVKAGVDIALMPASVTSLK EEEQKFARVIQALKEAVKNGDIPEQQINNSVERIISLKI KRG 397

10 Query: 357 LHNYEGRRQELFMPK-DKAMALINTLESQKIADEVADKAVTLVKDKQKDIFPVNPERYRH 415
 + Y R + K KA ++ + + K ++A+KAVT++K++Q + P P++
 Sbjct: 398 M--YPARNSDSTKEKIAKAKKIVGSKQHLKA EKKLA EKAVTVLKNEQHTL-PFKPKKGSR 454

15 Query: 416 ILLVNVEGYKGGFGAMIAGNKQRASDYMKE-----LL EAR GHEVTVWESTEERIMKLPQ 469
 IL+V + A +Q D +K L V+++ E+ +K
 Sbjct: 455 ILIV-----APYEEQTASIEQTIHDLIKRK KIKPVSLSKMNFA SQVFKTEHEKQVK--- 505

20 Query: 470 EERA AAIANVYAQK-QPIANL TEHYDLI INLV DVNAGGTTQR IIWPAAKGTPDQPFYVHE 528
 E I Y K P+ N D +I+ D + + ++P A + H
 Sbjct: 506 -EADYI ITGSYVVKNDPVVN----DGVID--DTISDSSKWATVF PRA--VMKAALQHN 554

25 Query: 529 IPSIVISVQHAFALADMPQVGTYINAYD-----GLPSTI SAVVAKLAGESEFTGVSP 580
 P + +S + + + A+ + I Y L I A V + G + + G P
 Sbjct: 555 KPFV LMSLRNPYDAANFEEAKALIAVYGF KG YANGRYLQPNI PAGVMAIFGQAKPKGTL P 614

Query: 581 VD 582
 VD
 Sbjct: 615 VD 616

No corresponding DNA sequence was identified in *S.pyogenes*.

30 A related GBS gene <SEQ ID 8565> and protein <SEQ ID 8566> were also identified. Analysis of this protein sequence reveals the following homology to a lipoprotein, with homology with the following sequences in the databases:

29.5/52.3% over 422aa

Bacillus subtilis

35 EGAD|20114| hypothetical 70.6 kd protein in feua 5'region precursor Insert characterized
 SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW INTERGENIC REGION
 PRECURSOR (ORF1). Insert

characterized

GP|1944006|dbj|BAA19499.1||AB002150 YbbD Insert characterized

40 GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature protein may
 be membrane-anchored and start at Cys-17. 17.5% identity
 over 354-aa overlap with Candida pelliculosa beta-glucosidase.; putative Insert
 characterized

GP|2632433|emb Insert characterized

45 ORF00431(367 - 1557 of 2388)
 EGAD|20114|BS0166(36 - 458 of 642) hypothetical 70.6 kd protein in feua 5'region precursor
 {Bacillus subtilis} SP|P40406|YBBD_BACSU HYPOTHETICAL 70.6 KDA LIPOPROTEIN IN FEUA-SIGW
 INTERGENIC RECTON PRECURSOR (ORF1). GP|1944006|dbj|BAA19499.1||AB002150 YbbD {Bacillus
 subtilis} GP|438455|gb|AAA64351.1||L19954 possible N-terminal signal sequence; mature
 protein may be membrane-anchored and start at Cys-17. 17.5% identity over 354-aa overlap
 with Candida pelliculosa beta-glucosidase.; putative {Bacillus subtilis} GP|2632433|emb
 %Match = 9.6

50 %Identity = 29.5 %Similarity = 52.2
 Matches = 119 Mismatches = 183 Conservative Sub.s = 92

114 144 174 204 234 264 294 324
 LMVGDSLGLAAA EQNGIAFYPVLVGKEVKSWEILREDIGEAFAKGQFEQQRKESINTFWANLDN**KG*AMTHLVDLT

60 MRPVFPLILSAVLFLSCFFGA
 10 20

354 384 414 426 456 486 528
 KKPFNLNQEAI E WIEK TIN EM TLDE KIGQLFF-----NMGASRSEEV LTDV LD RYHIAAV RY NRGS-----SSEIYDQ

-483-

SEQ ID 1236 (GBS50) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 8; MW 69.2kDa).

GBS50-His was purified as shown in Figure 192, lane 5.

The GBS50-His fusion product was purified (Figure 192, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 264), which confirmed that the protein is immunoaccessible on GBS bacteria.

40 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 379

A DNA sequence (GBSx0410) was identified in *S.agalactiae* <SEQ ID 1237> which encodes the amino acid sequence <SEQ ID 1238>. Analysis of this protein sequence reveals the following:

45 Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2266 (Affirmative) < succ>
bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 380

A DNA sequence (GBSx0411) was identified in *S.agalactiae* <SEQ ID 1239> which encodes the amino acid sequence <SEQ ID 1240>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
10 bacterial cytoplasm --- Certainty=0.2279(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9705> which encodes amino acid sequence <SEQ ID 9706> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP: AAC21726 GB: U32690 oxidoreductase [Haemophilus influenzae Rd]
  Identities = 197/271 (72%), Positives = 229/271 (83%)

20 Query: 26 NKVVVITGAGGVLCGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVEEGGIAKAYKANVL 85
          NK+++ITGAGGVLC ++AK+ A A +ALLDLN EAA A EI + GG AKAYK NVL
  Sbjct: 15 NKLIIITGAGGVLCFLAKQLAYTKANIALLDDLNFEAADKVAKEINQSGGAKAKAYKTNVL 74

25 Query: 86 SKENLVEEVHQAVLEDLGPTDILVNGAGGNPKATTDNFHELDLPSETKTFELDEAGIS 145
          EN++EV + D G DIL+NGAGGNPKATTDNFHF+ DL T+TFF+LD++GI
  Sbjct: 75 ELENIKEVRNQIETDFGTCDILINGAGGNNPKATTDNFHFQFDLNETTRTFDLDKGIE 134

30 Query: 146 FVFNLNYLGLPTQVFAQDMVGREGANIINISSMNAFTPPLTKIPAYSGAKAAISNFTQW 205
          FVFNLNYLGLPTQVFA+DM+G++GANIINISSMNAFTPPLTKIPAYSGAKAAISNFTQW
  Sbjct: 135 FVFNLNYLGSLLPTQVFAKDMLGKQGANIINISSMNAFTPPLTKIPAYSGAKAAISNFTQW 194

35 Query: 206 LAVHFSKVGIRCNIAIPAGFLVTQNQRSLLFTEGDQPTARAEKILNNTPMGRFGEASELIG 265
          LAV+FSKVGIRCNIAIPAGFLV+NQN +LLF +G+PT RA KIL NTPMGRFGE+ EL+G
  Sbjct: 195 LAVYFSKVGIRCNIAIPAGFLVSQNQLALLFDTEGKPTDRANKILTNTPMGRFGESEELLG 254

Query: 266 GLFFLADEKSSSFVNGVVLPIDGGFAAYSGV 296
          L FL DE S+FVNGVVL+DGFF+AYSGV
  Sbjct: 255 ALLFLIDENYSAFVNGVVI.PVDGGFSAYSGV 285
```

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1241> which encodes the amino acid sequence <SEQ ID 1242>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0358(Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

50 An alignment of the GAS and GBS proteins is shown below:

```
Identities = 77/279 (27%), Positives = 125/279 (44%), Gaps = 19/279 (6%)

Query: 18 MSKTITFTNKVVVITGAGGVLCGYMAKEFAKAGAKVALLDLNQEAAQTFADEIVEEGGIA 77
          M + K+ +ITGA + +AK +A+AGA + D+ QE E G A
  Sbjct: 1 MENMFSLQKIALITGASYGIGFEIAKAYAQAGATIVFNDIKQELVDKGLAAYRELGIEA 60

Query: 78 KAYKANVLSKENLEEVHQAVLEDLGPTDILVNGAGGNPKATTDNFHELDLPSETKTF 137
```

-485-

Y +V + +++ + +++G DILVN AG
 Sbjct: 61 HGYVCDVTDEAGIQQMVSQIEDEVGAIDILVNNAG-----IIRTPML 103

5 Query: 138 ELDEAGISFVFNLNYLGTLLPTQVFAQDMVGREGANIINISSMNAFTPLTKIPAYSGAKA 197
 E+ V +++ + ++ M+ + IINI SM + + AY+ AK
 Sbjct: 104 EMAAEDFRQVIDIDLAPFIVSKAVLPSMIAKGHGKIIINICSMMSLGRETVSAYAAKG 163

10 Query: 198 AISNFTQWLAVHFSKVGCIRCNIAIPGFLVTNQNRSLLFTE-DGQPTARAEKILNNTPMGR 256
 + T+ +A F + I+CN I PG++ T Q L + DG + I+ TP R
 Sbjct: 164 GLKMLTKNIASEFGEANIQCNGIGPGYIATPQTAPLRERQADGSRHPFDQFIIAKTPAAR 223

15 Query: 257 FGEASELIGGLFFLADEKSSSFVNVVLPIDGGFAAYSG 295
 +G +L G FLA + +S+FVNG +L +DGG AY G
 Sbjct: 224 WGTTEDLAGPAVFLASD-ASN FVN GHILYVDGGILAYIG 261

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 381

A DNA sequence (GBSx0412) was identified in *S.agalactiae* <SEQ ID 1243> which encodes the amino acid sequence <SEQ ID 1244>. This protein is predicted to be D-mannonate dehydrolase (uxuA). Analysis of this protein sequence reveals the following:

Possible site: 42
 >>> Seems to have no N-terminal signal sequence
 25 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.3188 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

30 The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04425 GB:AP001509 D-mannonate dehydrolase [Bacillus halodurans]
 Identities = 202/343 (58%), Positives = 261/343 (75%)

35 Query: 1 MEMSFRWYGEDDPVTLENIGQIPTMKGIVTAIYDVPGEVWSRERIQQLKEKVEAAGLKI 60
 M ++ RW+G D V LE I QIP MKGIV+AIYDV VG VW +E+I LK +E GL +
 Sbjct: 1 MRLTMRWFGPSDKVKLEYIKQIPGMKGIVSAIYDVAVGGWPKEKILALKNNIERHGLTL 60

Query: 61 SVIESVPVHEDIKLGRPTRDILLIDNYIQTVKNLAAEGIDTICYNFMPVFDWTRTDLAYQY 120
 VIESVPVHEDIKLG+PTRD I+NY QT+++LA GIDT+CYNFMPVFDWTR+ L ++
 40 Sbjct: 61 DVIESVPVHEDIKLGPTRDRYIENYKQTLRHLAECGIDTVCYNFMPVFDWTRSQLDFKL 120

Query: 121 PDGSTALIFDETVSKKMDPVNGEELSLPGWDASYSKEEMKAIMDAYAEIDEKLWENLYF 180
 DGS ALI++E V + +P++GEL LPGWD SY E +K ++ AY +I EE LW++LYTF
 Sbjct: 121 EDGSEALIYEEDVISRTNPNSGELELPGWDTSYENESLKGVLQAYKKISEEDLWDHLYF 180

45 Query: 181 IKRIIPEAEAVGVKMAIHPPDDPPSIFGLPRIITGLEAIERFVKLYDSKSNGITLCVGSY 240
 ++ I+P A+ VG+KMAIHPPDDPP+SIFGLPRI+T +ER + LYDS ++GIT+C GS
 Sbjct: 181 VQAIMPVADEVGKMAIHPPDDPPSIFGLPRIITNKANLERLLSLYDSPNHGKITMCSSL 240

50 Query: 241 ASDPQNDVLEISRRAFELDRVNFVHARNIKLGDGKSFKESEAHPSSEYGSIDMYEVIKLCHE 300
 ++ ND+ E+ R R++F HARNIK +SF+ESAH SE GS++M ++K H+
 Sbjct: 241 GANEANDLPEMIRHFQQGRIHFAHARNIKRTGPRSFQESAHLSEAGSVNMVAMLKAYHD 300

55 Query: 301 FGFGEGAIRPDHGRMIWGETGRPGYGLYDRALGATYVSGLYEAV 343
 GF G +RPDHGRMIWGE GRPGYGLYDRALGATY++G++EAV
 Sbjct: 301 IGFTGPLRDPDHGRMIWGEKGRPGYGLYDRALGATYLNIGIWEAV 343

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 382

A DNA sequence (GBSx0413) was identified in *S.agalactiae* <SEQ ID 1245> which encodes the amino acid sequence <SEQ ID 1246>. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.2447 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 383

A DNA sequence (GBSx0414) was identified in *S.agalactiae* <SEQ ID 1247> which encodes the amino acid sequence <SEQ ID 1248>. This protein is predicted to be uronate isomerase. Analysis of this protein sequence reveals the following:

Possible site: 14
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3066 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

>GP:BAB04424 GB:AP001509 uronate isomerase [Bacillus halodurans]
 Identities = 215/465 (46%), Positives = 294/465 (62%), Gaps = 7/465 (1%)

Query: 3 FNTTETFMLKNQAAIQLYEE-VKROPIFDYHCHLDPKDIFEDHIFFDNIVDILWLGHDHYKWR 61
 F +E F+L N+ +LY K PI DYHCHL P++I+E+ F+N+ WLGGDHYKWR
 Sbjct: 4 FLSEDFLLMNEYDRELYYTFAKNMPICDYHCHLSPQEIWENKPENMTKAWLGGDHYKWR 63

Query: 62 LMRANGISEAEITGPASNLEKFKAFTLERAYGNPVYHWSAMELKVNFGVNEILTESNA 121
 MR NG+ E ITG A + EKF A+A+T+ + GNP+YHW+ MELK F ++ L E+N
 Sbjct: 64 AMRLNGVREEFITGGAPDKEKFLAWAKTVPKTIGNPLYHWTHMELKTYFFHQPLDETN 123

Query: 122 EEIYHRLNHFLKEHKISPRRLIADSKVMFIGTTDHPPLDTLEWHKKLAADESFKTVVAPTF 181
 E ++ N L++ +PR LI S V IGTID P D+L +H+KL AD++F V PTF
 Sbjct: 124 ENWDACNRLLQQEAFTPRALIERSNVRAIGTTDDPTDSLLYHQQLQADDTFHVKVIPTF 183

Query: 182 RPDEAF-IEHRHFVDFITKLGDITQKEITDFSTFIAAMEERIAYFAQNGCRASDISFTEI 240
 RPD A IE F D++ KL D+T + + F+ A++ER+ +F ++GCR+SD TE+
 Sbjct: 184 RPDGALKIEQDSFADWVAKLSDVTGESLDTLDAFLHALKERLTFDEHGCRSSDHDMTEV 243

Query: 241 VFEQTDEELNDLFNKVCCEGYIPNQSEISKWQTAVFMELCRLYKKYGFVTQVHFGALRNN 300
 F + +E E +F K + E K++T + L + Y G+V Q H G +RNN
 Sbjct: 244 PFVEVNEQEAQHIFRKRLANEGLTKVENEKYKTFMLTWLGKEYAARGWVMQWHIGVMRNN 303

Query: 301 HSTIFEKLGADVGVDLGD-QVALTVNMNRLLDSLVKKDSLPMIWIYNLNPAYNIAVANT 359

-487-

+S + KLG D G DS+GD Q+A +LLD L K+ +LPK I Y +NP N +A+
 Sbjct: 304 NSRMLHKLGPDTGFDIGDGQIAHAT--AKLLDDKQGALPKTILYCVNPNANYILASM 361

5 Query: 360 LANFQANELGVRSYLQFGAGWWFADTKLGMISQMNALAEQGMLANFIGMLTDSRSFLSYQ 419
 + NF E GVR +QFG+ WWF D GM Q+ LA G+L+NFIGMLTDSRSFLSY
 Sbjct: 362 IGNF--TBSGVRGKVQFGSAWWFNDHIDGMRRQLTDLASVGLLSNFIGMLTDSRSFLSYP 419

10 Query: 420 RHDYFRRILCTYLGEWIEEGEVPEDYQALGSMAKDIAYQNAVNYF 464
 RHDYFRRILC +G WI+EG++P D + G + +DI Y N V+YF
 Sbjct: 420 RHDYFRRILCQLIGSWIKEGQLPPDMERWGQIVQDICYNNVVDFY 464

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

15 Example 384

A DNA sequence (GBSx0415) was identified in *S.agalactiae* <SEQ ID 1249> which encodes the amino acid sequence <SEQ ID 1250>. This protein is predicted to be 2-dehydro-3-deoxyphosphogluconate aldolase/4-hydroxy-2-oxoglutarate al. Analysis of this protein sequence reveals the following:

Possible site: 43

20 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

bacterial cytoplasm --- Certainty=0.3883 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 25 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9703> which encodes amino acid sequence <SEQ ID 9704> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

30 >GP:AAD35160 GB:AE001693 2-dehydro-3-deoxyphosphogluconate
 aldolase/4-hydroxy-2-oxoglutarate aldolase [Thermotoga maritima]
 Identities = 93/199 (46%), Positives = 125/199 (62%), Gaps = 6/199 (3%)

35 Query: 37 KNNYFFAVIRGKSSEDALEIAKHAILEGGIRNIEVTFSTPEASKVIKQLSDDFKNNKEIIV 96
 K + AV+R S E+A E A GG+ IE+TF+ P+A VIK+LS F K I+
 Sbjct: 8 KKHKIVAVLTRANSVEEAKEKALAVFEGGVHLIEITFTVPDADTVIKELS--FLKEKGAI 65

40 Query: 97 GAGTVMTTELAKEAIDAGAKFLVSPHFDSDIANLANENKVYFPGCATATEIVVARKYKC 156
 GAGTV + E ++A+++GA+F+VSPH D +I+ E V+Y PG T TE+V A K
 Sbjct: 66 GAGTVTSVEQCRKAVESGAEFIVSPHLDEEISQFCKEKGVFYMPGVMTPTELVKAMKLGH 125

45 Query: 157 QIIKLFPGGVVGPGFIKDIHGPIPDVDLMPGGVSNSVVERKAGAVAVGVGSALSSKV 216
 I+KLFPG VVGP F+K + GP P+V +P+GGV++ NV EW KAG +AVGVGSAL
 Sbjct: 126 TILKLFPGEVVGPGQFVKAMKGPFPNVKFVPTGGVNLDNVCEWFKAGVLAVGVGSALVKGT 185

Query: 217 ATEGYDSVTKIAKQFVSAL 235
 D V + AK FV +
 Sbjct: 186 P----DEVREKAKAFWEKI 200

50 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1251> which encodes the amino acid sequence <SEQ ID 1252>. Analysis of this protein sequence reveals the following:

Possible site: 29

>>> Seems to have no N-terminal signal sequence

55 ----- Final Results -----
 bacterial cytoplasm --- Certainty=0.1039 (Affirmative) < succ>

-488-

```
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

5 Identities = 82/204 (40%), Positives = 132/204 (64%)

```
Query: 32 MLNQLKNNYFFAVIRGKSSADEIAKHAILEGGIRNIEVTFSTPEASKVIKQLSDDFKNN 91
       +L +LK N V+RG+SSE+AL + +I GGI+ IEVT++ P AS+VI QL++ FK +
Sbjct: 6  ILTKLKANRLVLVVRGESSEEALACSLASIEGGIKTIEVTYTNPFASEVIGQLAERFKED 65
10
Query: 92 KEIIVGAGTVMTTELAKEAIDAGAKFLVSPHFDSIANLANENKVVYFPGCATATEIVVA 151
       E+++GAGTV+ A++AI AGA+F+V P+F+ +A + + + Y PGC T E+V A
Sbjct: 66 PEVLIGAGTVLDDVTARQAILAGAQFTIVGPNFNRVALICHRYSIPIYLPGCCMTVNEVVTA 125
15
Query: 152 RKYKCQIILKFPGGVVPGFIKDIHGPIDVDLMPSGGVSVSNVVEWRKAGAVAVVGSA 211
       + ++K+FPG VG FI+ I P+P V++M +GGVS N+ +W AG +G+G
Sbjct: 126 LESGVDMVKIFPGSTVGISFIRAIKSPLPQVEVMVTGGVSSDNLKDWLAAAGDVVLGIGGE 185
20
Query: 212 LSSKVATEGYDSVTKIAKQFVSAL 235
       + + + Y+ +TK A ++ +L
Sbjct: 186 FNQLASQKQYNLITKKAHYIKSL 209
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

25 Example 385

A DNA sequence (GBSx0416) was identified in *S.agalactiae* <SEQ ID 1253> which encodes the amino acid sequence <SEQ ID 1254>. This protein is predicted to be pyruvate dehydrogenase complex repressor. Analysis of this protein sequence reveals the following:

30 Possible site: 26
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

```
bacterial cytoplasm --- Certainty=0.2827(Affirmative) < succ>
bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
35        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12044 GB:Z99105 similar to transcriptional regulator (GntR
family) [Bacillus subtilis]
40        Identities = 67/225 (29%), Positives = 119/225 (52%), Gaps = 17/225 (7%)
```

```
Query: 3 RPLVEQTADRLLLHLIEREYPVGA KLPNEYELAEDLDVGRSTIREAVRSLATRNILEVRQ 62
       + L +Q +R++HL+ + G KLP E EL + L V R +REA+ SL T ++ +
Sbjct: 16 KTLAKQVIERIVHLLSSGQLRAGDKLPTEMELMDILHVSRPVLREALSSLTLGVITRK 75
45
Query: 63 GSGTYISSLKGVS DPLGFSLIKDTDRLLTSDLFELRLLEPRIAELVAYRITDDQLQLLE 122
       GTY + K G+ P L TD L + + E R+ LE + + A +I +++LQ L+
Sbjct: 76 RG GTYFNDKIGM--QPF SVM LALATDNLP A-II EARM AELGLVTIAAEKINEELQR LQ 132
50
Query: 123 KLVGDIEDAV--HAGDPKHL LLDVEFHSM LAKYSGNIAMDSLLPVINQSIHLINANYTNR 180
       K + DI ++ H G+ D EFH ++A + N ++ ++ QS+ + +A ++
Sbjct: 133 KTIDDIA STDNHYGE----ADKEFHR II A L S A NNP V VEG M I---QSLLI THAKIDSQ 183
55
Query: 181 ---QMKS D SLEAHREIIKAIREKNPVAHDAMLMHIMS VRRSALK 222
       + + ++E H++I A+ +++P AH M H+ VR LK
Sbjct: 184 IPYRERDVT VEYHKKIYDALAQRDPYKAHYHMYEHLKFVRDKILK 228
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1255> which encodes the amino acid sequence <SEQ ID 1256>. Analysis of this protein sequence reveals the following:

-489-

Possible site: 54

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

5 bacterial cytoplasm --- Certainty=0.2161(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

10 Identities = 24/51 (47%), Positives = 35/51 (68%)

Query: 22 YPVGAKLPEYELAEDLDVGRSTIREAVRSLATRNILEVRQGSGTYISSKK 72
 +P+G++L+P+E LAE V R T+R+A+ L ILE R GSGTY++S +
 Sbjct: 30 WPIGSRLPSERHIAEHFTVSRMTLRQAITLLVEEGILERRIGSGTYVASHR 80

15 Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 386

A DNA sequence (GBSx0417) was identified in *S.agalactiae* <SEQ ID 1257> which encodes the amino acid sequence <SEQ ID 1258>. Analysis of this protein sequence reveals the following:

Possible site: 31

>>> Seems to have no N-terminal signal sequence

----- Final Results -----

25 bacterial cytoplasm --- Certainty=0.2178(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

A related GBS nucleic acid sequence <SEQ ID 9701> which encodes amino acid sequence <SEQ ID 9702> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

>GP:CAA58911 GB:X84105 glucuronidase [synthetic construct]
 Identities = 258/602 (42%), Positives = 357/602 (58%), Gaps = 31/602 (5%)

35 Query: 23 MLYPLLTKTRNTYDLGGIWNFKLGEHNP-----ELLPSPDEVMDVIPTSFNDLMVSKEK 75
 ML P+ T TR L G+W F L N L + +P SFND +
 Sbjct: 1 MLRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQESRAIAVPGSFNDQFADADI 60

40 Query: 76 RDYIGDFWYEKVIEWPKVSEDEEMVLRFGSVTHQAKIYVDGVLVGHEHKGGFTPFEVLVPE 135
 R+Y G+ WY++ + +PK + +VLRF +VTH K++V+ V EH+GG+TPFE V
 Sbjct: 61 RNYAGNWWYQREVFIPKGWAGQRIVLRFDAVTHYGKVVNNQEVMEHQGGYTPFEADVTP 120

45 Query: 136 CKYNNEKIKVSIICANNVLDYTTLPGVNYSEIIIQEDGSIKKKVRENFDFFNYAGVHRPLKL 195
 + +++++C NN L++ T+P G I E+G KKK DFFNYAG+HR + L
 Sbjct: 121 YVIAGKSVRITVCVNNELNWQTIPPGMV--ITDENG--KKQSYFHDDFNYAGIHRSQL 176

50 Query: 196 MIRPKNHIFDITITSRLSDLQSAIDLHFLVETTNOKVDEVRISVFDEDNKLV--GETKDSR 253
 P + DIT+ + ++ D A + + V N +V + + D D ++V G+
 Sbjct: 177 YTTPNITWVDDITVVTHVAQDCNHASVDWQVAN--GDVSVELRDADQQVATGQGTSGT 233

55 Query: 254 LFLSDVHLWEVLNAYLYTARVEIFVDNQLQDVYEEFGLREIEVTNGQFLINRKPIYFKG 313
 L + + HLW+ YLY V + D+Y G+R + V QFL+N KP YF G
 Sbjct: 234 LQVVNPQLWQPGEGYLYELCVTAKSQTEC-DIYPLRVGIRSAVKGEQFLINHKPFYFTG 292

55 Query: 314 FGKHEDTFINGRGLNEAANLMDILNLKDMGANSFRSHYPYSEEMMRLADRMGVLVIDEV 373
 FG+HED + G+G + + D L+ +GANS+RTSHYPY+EEM+ AD G++VIDE
 Sbjct: 293 FGRHEDADLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYAEEMLDWADEHGVVIDET 352

Query: 374 PAVGLFQNFNASLDLS-----PKDNGTWNL--QTKAAHEQAIQELVKRDKNHPSSVMW 425

-490-

Avg FN SL + PK+ + + +T+ AH QAI+EL+ RDKNHPSVVMW
 Sbjct: 353 AAVG----FNLSLGIGFEAGNPKELYSEEAVNGETQQAHLQAIKELIARDKNHPSVVMW 408

5 Query: 426 VVANE PASHEAGAH DYFEPLVKLYKDLD PQKR PVT LVN ILMAT PDRD QVMDL DVV CLNR 485
 +ANEP + GA +YF PL + + LDP RP+T VN++ D + DL DV+CLNR
 Sbjct: 409 SIANEPDTRPQGAREYFAPLA EATRKLDPT-RPITCVNMFCDAHTDTISDLFDVLCLNR 467

10 Query: 486 YYG WYV DHD GLT NAEVGIR KELLE WQDKFP DKPIII TEYGA DTLPGL HSTWNIP YTEE FQ 545
 YYG WYV GDL AE + KELL WQ+K +PIII TEY G DTL GLHS + ++EE+Q
 Sbjct: 468 YYG WYV QSGD LETAE KVLE KELL AWQE KL-HQPIII TEYGV DTL AGL HSMY TDMW SEE YQ 526

15 Query: 546 CDF YEM SHRV FDG IPNL VGEQV WNFAD FETN LMIL RVQGN HKGL FSRN RQPK QVV KEF KK 605
 C + +M HRV FD + +V GEQV WNFAD F T+ ILRV GN KG+F+R+R+PK +K
 Sbjct: 527 CAW LD MYH RVF DRV SAV VGEQV WNFAD FATS QGIL RVGGN KKG I FTR DRK PKS A FLL QK 586

Query: 606 RW 607
 RW
 Sbjct: 587 RW 588

20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1259> which encodes the amino acid sequence <SEQ ID 1260>. Analysis of this protein sequence reveals the following:

Possible site: 23

25 >>> Seems to have no N-terminal signal sequence
 INTEGRAL Likelihood = -4.04 Transmembrane 1131 -1147 (1130 -1147)

30 ----- Final Results -----
 bacterial membrane --- Certainty=0.2614(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the databases:

>GP:AAF97242 GB:AF282987 beta-galactosidase precursor [Streptococcus pneumoniae]
 Identities = 303/921 (32%), Positives = 463/921 (49%), Gaps = 86/921 (9%)

35 Query: 5 QKSSE I VT---RTITKPSRATSNVKQEI DMTPDSKE QT V TGYQYHYIDQ--EGRK QPFN 58
 +K E VT + KP ++ + ++ + +Q E RK FN
 Sbjct: 96 KKE DEAVTPK EKVSAKPEEKAPRIESQASNQEKPLKEDAKAVTNEEVNQMIEDRKVDFN 155

40 Query: 59 QGWRF-LMADVACAQDPSFDDSNWQVIHLPHDFSLTQPYTRNGEA--ESAYKLGGVGWYR 115
 Q W F L A+ A P D S W+ + LP+D+S+ + A E GG WYR
 Sbjct: 156 QNWYFKLNANSKEAIKP DAD VSTWKLDL PYDWSI FND FDHESPAQN EGGQLNGGEAWYR 215

45 Query: 116 HYLV LDEV LAGCHVA ITFEGSY METE IYV NGQ FIGKHL NGYQEFTYD ISDV VTF-GAENL 174
 LDE +V +TP+G YM+++++YVNGQ +G + NGY +F+YDI+ + G EN+
 Sbjct: 216 KTFKLDEKDLKKNVRLTFDG VYMD S QVYV NGQLV GHYPNGYNQFSYDITKYLQKD GREN V 275

50 Query: 175 LAVR VEN KVPSS RWYSGSG LYREV SLS VL PQL HFVAD QVAM TLAD TAVQ EKG QQK VDL RF 234
 +AV NK PSS RWYSGSG +YR+V+L V ++H + + Q+ G+ + +
 Sbjct: 276 IAV HAV N KQPSS RWYSGSG IYRD VTL QVTD KVHV EKNGTT ILTPK LEEQ QHG KVETH VTS 335

55 Query: 235 ALNQS IQTCHY QLSCL W EQSHCS KDKL L YQETEVPLAD IAFQRQYGLT--LSLEELQL 292
 + + H ++ E + + + L L L +E + L
 Sbjct: 336 KIVNT DDKDH ELV A---EYQ IVERGGH A VTL QVTD KVHV EKNGTT ILTPK LEEQ QHG KVETH VTS 391

Query: 293 WSP--DNPHLYDLELTLYYQGQVIDCFCL ETGFRQLTFMANQGLFVN GRAV KLG V CLHH 350
 W+ D P LY+L +Y GQ+ +D G+R + N+G +NG +K GV LHH
 Sbjct: 392 WT VLN DKP ALYEL ITRVY RDG QL VDA KDL FG YR YYHWT PNEG FSLN GERIK F HG VSL HH 451

60 Query: 351 DQGG L GACAYE DALAROL VLL KDMGANTIR STHN PSSPK LRQ LAN RL GFVIEEA FDTWT 410
 D G LGA A R+L +K+MG N+IR+THNP+S + Q+A LG V EEA FDTW
 Sbjct: 452 DHG ALGA EEN YKA EYR RL KQM KEMGV NSIR T HNPASE QTL QIAA E L GLL VQEE AFDTWY 511

65 Query: 411 YAKNG NVN DFS NYF HQTIGTEN ANYL QRV RSP ET SWA QYS IEAM VWS A KND P S VLM WSIG 470
 K D+ +F+ A ++ W+ + + MV KN+P++ MWSIG

-491-

Sbjct: 512 GGK--KPYDYGRFFEKDATHPEARKGEK-----WSDFDLRTMVERGKNNPAIFMWSIG 562

Query: 471 NELMEGFSADVSHYPELTROMCQWITAIDTSRPITFGDNLKEADFC-WHEEVSQMATLL 529

NE+ G + +H +++ + I +D +R +T G +K + + HE+++

5 Sbjct: 563 NEI--GEANGDAHSLATVKRLVKVIKDVDKTRYVTMGADKFRFGNGSGGHEKIA----- 614

Query: 530 SQLDHPQGLIGLNYADGKDYDRLHEEHSDWLLYGSETVSAITSR-AYYKETKKVLDs--- 585
+LD +G NY++ +Y L +H WL+YGSET SA +R +YY+ +++ S

10 Sbjct: 615 DELD----AVGFNYSE-DNYKALRAKHPKWLTYGSETSSATRTRGSYYRPERELKHSNGP 669

Query: 586 --GYHLTSYDHAKVDWGAFAQSQAWYDTITRDFV--AGECVWTGFDYLGEPTPWNKIDSGV 641
Y + Y + +V WG A+ +W T RD AG+ +WIG DY+GEPTPW+ +

15 Sbjct: 670 ERNYEQSDYGNDRVGWGKTATASW--TFDRDNAGYAGQFIIWTGTDYIGEPTPWHNQNQTP 727

Query: 642 VGLWPSPKNAYFGIILDTAGFPKDSYYFYQSOW--AQGQTTLHLLPVWQKD----QLCFD 694
V K++YFGI+DTAG PK +Y YQSOW + + +HLLP W + D

Sbjct: 728 V-----KSSYFGIVDTAGIPKHDFYLYQSOWSVVKKKPMVHLLPHWNWENKELASKVAD 781

20 Query: 695 EQGLVVEVVVVSNAASVQILMFEDEQGNLTDYGRKAFHTYSTPTGHTYQLYQGADAAKNPHE 754
+G + V YSNA+SV+L N G K F+ T G TYQ +GA+A

Sbjct: 782 SEGKIPVRAYSNASSVELFL-----NGKSLGLKTFNKKQTSDGRTYQ--EGANA-----N 829

25 Query: 755 NLYLTWRVPYQKGLLRAVAYDISGKSIPKTSGRSQVRTYGSVAKLSWKAFEAPIDAPW-E 813
LYL W+V YQ G L A+A D SGK I R ++ T G A + + IA +

Sbjct: 830 ELYLEWKVAYQPGTLEATARDGESGKEI---ARDKITTAGPKAAVRLIKEHDHAIADGKD 885

Query: 814 LLYLDLSSLDSRGELVLSHAQDWLQVQVEGPARRLLADNGNPTDHTPYQEP----LRQAY 868
L Y+ ++DS+G +V A + ++ Q+ G +L+ +DNG Y+ +R+A+

30 Sbjct: 886 LTYIYYEIVDSQGNVVTANNLVRQLHGQGQLVGVDNQEASRERYKAQADGSWIRKAF 945

Query: 869 GGKLLAILALTGEAGHIKVTA 889

GK +AI+ T +AG +TA

Sbjct: 946 NGKGVAIVKSTEQAGKFTLTA 966

35

An alignment of the GAS and GBS proteins is shown below:

Identities = 98/414 (23%), Positives = 175/414 (41%), Gaps = 64/414 (15%)

40 Query: 54 LPSDEVMDVIPTSFNDLMSKEKRDYIGDFWYEKVIEWPKVSEDEEMVLRFGSVTHQAKIY 113
LP D + P + N S K +G WY + + +V + + F + +IY

Sbjct: 86 LPHIDFSLTQPYTRNGEAESAYKLGGVG---WYRHYLVLDEVLAGCHVAITPEGSYMETIY 143

Query: 114 VDGVLVGEHKGGFTPFEVLVPECKYNNEKIKVSIICANNVLDYTTLPVGNYSEIIQEDGSI 173
V+G +G+H G+ F + + V+ A N+L +

45

Sbjct: 144 VNGQFIGKHLNGYQEFTYDISDV-----VTFGAENLLAVR-----V 179

Query: 174 KKKVRENFDFFNYAGVHRLPLKLMIRPKNHIFDITITSRLSDL-----QSADLHFLVET 227
+ KV + + + +G++R + L + P+ H + L+D Q DL F +

50

Sbjct: 180 ENKVPSS-RWYSGSGLYREVSLVLPQLHFVADQVAMTLADTAVQEKGQQKVDLRFALMQ 238

Query: 228 NQKVDENVRIASFV-----DEDNKLVGETKDS-----RLFLSDVHLWEVLNA 267
+ + +S+ +D KL+ + + L L ++ LW N

Sbjct: 239 SIQTCHYQLSLCLWEQSHCSKDKLKYQETEVPLADLAFAQRQYGLTLSLEELQLWSPDNP 298

55

Query: 268 YLYTARVEIFVNDLNQLQDVYEEFNGLRETE-VTNGQFLINRKPIYFKGFGKHEDTFINGRG 326
+LY + ++ Q+ D + G R++ + N + + KG H D G

Sbjct: 299 HLYDLELTLYQQVIDCFCLETGFRQLTFMANQGLFVNNGRAVKLKGVLHHDQGGLGAC 358

60

Query: 327 LNEAANLMDLNLLKDMGANSFRTSHYPSEEMMRЛАDRMGVLVIDEPAVGLFQ---NEN 383
E A L LLKDMGAN+ R++H P S ++ +LA+R+G VI+E + N N

Sbjct: 359 AYEDALARQLVLLKDMGANTIRSTHNPSSPKLRLQLANRLGFFVIEEAFDTWTYAKNGNVN 418

Query: 384 ASLDLSPKDNGTWN---LMQTKAAH---EQAIQELVKRDKNHPSVVMVVANE 430
+ + GT N L + ++ + +I+ +V KN PSV+MW + NE

65

Sbjct: 419 DFSNYFHQTIGTENANYLQRVRSPETSWAQYSIEAMVWSAKNDPSVLMWSIGNE 472

-492-

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 387

A DNA sequence (GBSx0418) was identified in *S.agalactiae* <SEQ ID 1261> which encodes the amino acid sequence <SEQ ID 1262>. This protein is predicted to be 2-keto-3-deoxygluconate kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -0.53    Transmembrane 197 - 213 ( 197 - 213)

10      ----- Final Results -----
      bacterial membrane --- Certainty=0.1213 (Affirmative) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
      bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

15 A related GBS nucleic acid sequence <SEQ ID 9699> which encodes amino acid sequence <SEQ ID 9700> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD35161 GB:AE001693 2-keto-3-deoxygluconate kinase [Thermotoga maritima]
  Identities = 115/342 (33%), Positives = 180/342 (52%), Gaps = 16/342 (4%)
```

```
Query: 14 KIISLGVEVLLRLSPPQYHTLMQANHLKCQFGGSELNVLASLAQLGYHVGLVSALPDNDLG 73
  K+++ GE++LRLSPP + + Q + +GG+E NV A LAQ+G V+ LP+N LG
```

```
25 Sbjct: 2 KVVTFGEIMRLSPPDHKRIFQTDSFDVTYGGAEANVAFLAQMGLDAYFVTKLPNNPLG 61
```

```
Query: 74 KMASQFILSQQISPAAIKKKEGRGLIYYYYEQQFSVRTNKVIYDRNYSSFWESTLSDYDFT 133
  A+ + + I + R+GIY+ E G S R +KV+YDR +S+ E+ D+D+
Sbjct: 62 DAAAGHRLRKGFGVKTDYIARGGNRIGIYFLEIGASQRPSKVYDRAHSAISEAKREDFDWE 121
```

```
30 Query: 134 SIFKGVDWFHVSGITPALTKDLYEVTRFLMTKAEGGVKVSIDLNFRESLWSSFQEAREQ 193
  I G WFH SGITP L K+L + + A E GV VS DLN+R LW+ +EA++
Sbjct: 122 KILDGARWFHFSGITPPLGKELPLILEDALKVANEKGTVSCDLNYRARLWTK-EEAQKV 180
```

```
35 Query: 194 LSPLLGLLDVCFGLEPIYLAGESEDLKDELGLSRPYLDI-----ELLEKITQKIVQEY 246
  + P + +DV L ED++ LG+S LD+ E KI +++ ++Y
Sbjct: 181 MIPFMNEYVDV-----LIANEEDIEKVLGVISVEGLDLKTGKLNREAYAKIAEEVTRKY 232
```

```
40 Query: 247 GLDYIAFTQREMEYTNQYMLKSYLYHNNMLYQIDKTGVEVLDRTVGTGDAFAAGLIHALLE 306
  + T RE + +N + +++ + ++DRVGD+FA LI+ L
Sbjct: 233 NFKTVGILTRESISATVNYWSVMVFENGQPHFSNRYEIHIVDRVGAGDSFAGALIYGSLM 292
```

```
45 Query: 307 KETPQRALEIAMATFKYKHTIQGDINIMTRDDIAYLIEKTN 348
  Q+ E A A KHTI GD +++ ++I L T+
Sbjct: 293 GFDSQKKAEEFAAAASCLKHТИPGDFVVLSIEEIKLASGATS 334
```

45 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1263> which encodes the amino acid sequence <SEQ ID 1264>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
50      ----- Final Results -----
      bacterial cytoplasm --- Certainty=0.0708 (Affirmative) < succ>
      bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
      bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

55 An alignment of the GAS and GBS proteins is shown below:

```
Identities = 111/319 (34%), Positives = 168/319 (51%), Gaps = 7/319 (2%)
```

Query: 12 MAKIISLGEVLLRLSPPQYHTLMQANHLKCQFGGSELNVLASLAQLGYHGLVSALPDND 71
M+K++ +GE L+R+SP Q+ L A + FGGSE+N+ +L G L +ALPDN ;
Sbjct: 14 MSKLLLVGEPLIRVSPNQFQPLTNACEAQLFFGGSEVNIAARTLGGFGLEARLFTALPDNP 73

5

Query: 72 LGKMASQFILSQQISPAAIIKKKEGRGLGIYYYYEQGFSVRTNKVIYDRNYSSFWESTLSDYD 131
+G QF+ + + + R+G+YY E GF R ++V YDR SSF D
Sbjct: 74 VGHAFHQFLKQSGVDMSLTAWQGHRVGLYYLENGFGCRASQVYYDRCGSSFSALDKDSL 133

10 Query: 132 FTSIFKGVDWFHVSGITPALTKDLYEVTFRFLMTKAEGGVKVKSIDLNFRFRESLWSSFQEARI 191
+IF+G+ FH SGI+ AL K ++ L+ +AK+ + +S DLNFR S+ + +A+
Sbjct: 134 LAAIFEGISHFHFGISLALGKKTQDLIEVLVREAKKRDICISFDLNFRSSM-IAVADAK 192

15 Query: 192 EQLSPLLGLLDVCFCGLEPIYLAGESEDLKDELGLSRPYLDIELLEKITQKIVQEYGLDYI 251
S D+ FG+EP+ L + D+ D R D + + + Q Y L I
Sbjct: 193 RLFSHFAQYADIIFGMEPLLLSDDFDMFD----RKKADTTTIRERLAGLYQRYQLQAI 247

20 Query: 252 AFTQREMEYTNQYMLKSYLYHNINMLYQTDKTVGVEVLDRVGTGDAFAAGLIHALLEKETPQ 311
T+R + K+Y Y + Y++ + VL RVG+GDAF AGL++ LLE Q
Sbjct: 248 YHTERSNDAQGSNHFKAYAY-DRQFYESCEVTPVLQRVGSGDAFVAGLLYQLLEGNEKQ 306

25 Query: 312 RALEIAMATFKYKHTIQGD 330
R L+ A+AT K T+ D
Sbjct: 307 RNLDFAVATASLKCTVAED 325

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 388

A DNA sequence (GBSx0419) was identified in *S.agalactiae* <SEQ ID 1265> which encodes the amino acid sequence <SEQ ID 1266>. Analysis of this protein sequence reveals the following:

Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
INTEGRAL Likelihood = -1.17 Transmembrane 5 - 21 (5 - 21)
35 ----- Final Results -----
bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

40 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 389

45 A DNA sequence (GBSx0420) was identified in *S.agalactiae* <SEQ ID 1267> which encodes the amino acid sequence <SEQ ID 1268>. Analysis of this protein sequence reveals the following:

Possible site: 53
>>> Seems to have no N-terminal signal sequence
INTEGRAL Likelihood = -12.05 Transmembrane 198 - 214 (191 - 220)
50 INTEGRAL Likelihood = -11.68 Transmembrane 446 - 462 (437 - 467)
INTEGRAL Likelihood = -9.55 Transmembrane 94 - 110 (91 - 116)
INTEGRAL Likelihood = -7.43 Transmembrane 291 - 307 (283 - 309)
INTEGRAL Likelihood = -4.88 Transmembrane 265 - 281 (257 - 282)
INTEGRAL Likelihood = -4.62 Transmembrane 321 - 337 (318 - 339)
55 INTEGRAL Likelihood = -3.93 Transmembrane 406 - 422 (405 - 426)
INTEGRAL Likelihood = -1.59 Transmembrane 121 - 137 (121 - 137)

-494-

INTEGRAL Likelihood = -1.12 Transmembrane 345 - 361 (345 - 362)
 INTEGRAL Likelihood = -0.48 Transmembrane 43 - 59 (43 - 59)

----- Final Results -----

5 bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

10 >GP:CAB13641 GB:Z99113 similar to H⁺-symporter [Bacillus subtilis]
 Identities = 105/452 (23%), Positives = 182/452 (40%), Gaps = 37/452 (8%)

15 Query: 36 IYLFTFMFVTYFSTGVLSAAIFVSQIMGYIRIFDGFIDPAIGIMIDKTDKFGKYRPI 95
 IY ++ +F T V G +A + +RI D DP IG ++D+T++F ++RP L
 Sbjct: 27 IYATVSTYLLFFYTDVFGLSAAAAGTMFLVVRIIDALADPFIGTIVDRTNSRFARFRPYL 86

Query: 96 IIIGNVITALSLIFLIALRGVDENIRFPLFILVLIIHKGYSQQTITKAGQTALTNDPKQ 155
 + G A + L L + ++ I +G S+ T ALT+
 Sbjct: 87 LFGL---AFPFVILAILCFTTPDFSDMGKLIYAYITYVGLSLTYTTINVPGALTS-AMT 141

20 Query: 156 RPIFNIVDAVMTTSIIMTGGQFVVSVFVLPKFGNFTPQFFNVLIFGTILISAILAIV--AI 213
 R +V L +V F VP + G L IL ++ +
 Sbjct: 142 RNNQEVSITSVRMLFANLGGLVVAFFVPLLAAYLSDTSGNESLGWQLTMGILGMIGGCL 201

25 Query: 214 IGIWAKDRKEFFGLGENTQKTALKDYWKVLGNKPLQILSIAALVKFAIQFFGDSV-VM 272
 + K KE L ++ +K D ++ + N+PL +LSI ++ F + +SV +
 Sbjct: 202 LIFCFKSTKERVTLQKSEEKIKFTDIFEQFRVNRLPVLVLSIFFIII-FGVNSISNSVGIY 260

30 Query: 273 VLLFGI---LFGNYALSGQFSLLFIVPGVIINILFSTIARKKGLRFSYVRAIQIGMIGL 328
 + + + L Y L G L I+P I L + +KK L + A+ + +IGL
 Sbjct: 261 YVTYNLEREDLVWKYGLIGSLPALVILP--FIPRLHQFLGKKLLNY---ALLNIIGL 314

35 Query: 329 LAFGAVLYVGKPGDSLTLNLYYTILFIVTNITIARYASQAPASLVLTMGADISDYETSES 388
 LA L + N+Y IL V +IA S + + + +Y +
 Sbjct: 315 LAL-----LFVPPSNVYLIL--VCRLIAAAAGSLTAGGYMWALIPIPTEYGEYRT 361

40 Query: 389 GRYVSGMIGTIFS LTDIASSFAPMVVGFLAGIGFSKSFPPTIETPLPPDLKMAAISILV 448
 G+ + G+I I + +V G VL G+ P M +
 Sbjct: 362 GKRMCGLIYAIIGFFFKEGMALGGVVPGLVLDKFGY----VANQAQTPAALMGILITTT 416

Query: 449 AIPFIALSIALLMKFYKLDKEEMVRIQEKIQ 480
 IP L +AL+ + FY LD+++ + +++
 Sbjct: 417 IIPVFLLVLALIDINFYNLDEKKYKNMVRLE 448

45 No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 390

A DNA sequence (GBSx0422) was identified in *S.agalactiae* <SEQ ID 1269> which encodes the amino acid sequence <SEQ ID 1270>. Analysis of this protein sequence reveals the following:

Possible site: 52
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

55 bacterial cytoplasm --- Certainty=0.3375(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

The protein has homology with the following sequences in the GENPEPT database:

-495-

>GP:AAB17663 GB:U31175 D-specific D-2-hydroxyacid dehydrogenase [S. aureus]
 Identities = 165/331 (49%), Positives = 231/331 (68%), Gaps = 1/331 (0%)

Query: 1 MMKLIKVNREEEATLAQDWANRNHVELSMSEGPLTLETVNEVEGFDGIANAQIEPLDDA 60
 5 M K+ F R+ E +A +W +N+VE++ S+ L+ TV++++ +DG+ Q L++
 Sbjct: 1 MTKIMFFGTRDYEKEMALNWGKKNNVEVTTSKELLSSATVDQLKDYDGVTTMQFGKLEND 60

Query: 61 IYPLLKEMGIKQIAQRSAGVDMYNILEAKQHGIIISNVPSYSPESTIAEFTVTIALNLIRK 120
 10 +YP L+ GIKQIAQR+AG DMY+L+LAK+H I+ISNVPSYSPE+IAE++V+IAL L+R+
 Sbjct: 61 VYPKLESYGIKQIAQRTAGFDMDYDLDLAKKHNIVISNVPSYSPESTIAEYSVSIALQLVRR 120

Query: 121 VELIRANVREQNFSWTLPIRGRVLGNMTVAIIGTGRIGLATAKIFKGFGCRVIGYDIYHN 180
 15 I V+ +F+W I + + NMTVAIIGTGRIG ATAKI+ GFG + YD Y N
 Sbjct: 121 FPDIERRVQAHDFTWQAEIMSKPVKNMTVAIIGTGRIGAATAKIYAGFGATITAYDAYPN 180

Query: 181 PMADGILEYVNSVVEEAVVEADLVLSPHMPPTAENTHLFNLDMFKQFKKGAILMNARGALV 240
 20 D L Y +SV+EA+++AD++SLH+P E+ HLF+ MF KKGAIL+N ARGA++
 Sbjct: 181 KDLD-FLTYKDSVKEAIKDADIISLHVPAKESYHLFDKAMFDHVKKGAILVNAARGAVI 239

Query: 241 ETKDLLEALDQGLLEGAGIDTYEFGPYIPKNCQGQDISDKDFLRLINHPKVIYTPHAAY 300
 25 T DL+ A++ G L GA IDTYE E Y + +DI DK L LI H +++ TPH A+
 Sbjct: 240 NTPDLIAAVNDGTLLGAAIDTYENEAAFTNDWTNKDIDDKTILLELIEHERILVTPHIAF 299

Query: 301 YTDEAVKNLVEGALNACVEVIETGTTTKVN 331
 25 ++DEAV+NLVEG LNA + VI TGT T++N
 Sbjct: 300 FSDEAVQNLVEGGLNAALSVINTGTCETRLN 330

There is also homology to SEQ ID 124.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 391

A DNA sequence (GBSx0423) was identified in *S.agalactiae* <SEQ ID 1271> which encodes the amino acid sequence <SEQ ID 1272>. Analysis of this protein sequence reveals the following:

Possible site: 40
 35 >>> Seems to have no N-terminal signal sequence
 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.2364 (Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000 (Not Clear) < succ>
 40 bacterial outside --- Certainty=0.0000 (Not Clear) < succ>

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 392

A DNA sequence (GBSx0424) was identified in *S.agalactiae* <SEQ ID 1273> which encodes the amino acid sequence <SEQ ID 1274>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

Possible site: 37
 50 >>> Seems to have a cleavable N-term signal seq.
 INTEGRAL Likelihood =-12.90 Transmembrane 64 - 80 (53 - 89)
 ----- Final Results -----

-496-

```
bacterial membrane --- Certainty=0.6158 (Affirmative) < succ>
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

- 5 A related GBS nucleic acid sequence <SEQ ID 9325> which encodes amino acid sequence <SEQ ID 9326> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema pallidum]
```

10 Identities = 33/91 (36%), Positives = 55/91 (60%), Gaps = 1/91 (1%)

```
Query: 1 MANVIAKPKIMLPMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGGWSV 60
M N + P + +P++ + + G+L LFN+QGTPASAGFG GL+GPINA L V
```

```
Sbjct: 250 MPNWIRYPILNIPLLLNLNGLVCGVLAFLFNLQGTPASAGFGFIGLVGPINAYRLMAYTPMV 309
```

15 Query: 61 MNMLLIIIIIFVAAPPIIILNFIFNYLFIKVLKI 91
+L ++ FV + + ++ +++ + LK+
Sbjct: 310 RAGILFLVYFVLS-FLAAYLIDFILVDRKLK 339

- 20 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1275> which encodes the amino acid sequence <SEQ ID 1276>. Analysis of this protein sequence reveals the following:

Possible site: 51

>>> Seems to have a cleavable N-term signal seq.

```
INTEGRAL Likelihood = -12.31 Transmembrane 141 - 157 ( 133 - 166)
25 INTEGRAL Likelihood = -6.00 Transmembrane 92 - 108 ( 88 - 112)
```

----- Final Results -----

```
bacterial membrane --- Certainty=0.5925 (Affirmative) < succ>
```

```
bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

```
bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema pallidum]
```

35 Identities = 63/178 (35%), Positives = 107/178 (59%), Gaps = 10/178 (5%)

```
Query: 2 IGQQIASLLGLQPILMSLLIAMIFCFILIVSPITTVGIALAINLSGIGSGAASFG----- 55
+G+ IA+ + LQP+LMS+L++M F +I+SP+++V + +A+ L+G+ SGAA+ G
```

```
Sbjct: 164 VGRVIATFIALQPLLMSILLSMSFSLIIISPVSSVAVGIAVGLTGLASGAANIGVSSCAM 223
```

40 Query: 56 -LCLAGWAVNSKGTSLAHVLRSPIKISMANVLSKPKIMLPMLCSAAVLGVIGAIFNIQGTP 114
L + VN G IA + K+ M N + P + +P+L + V GV+ +FN+QGTP
Sbjct: 224 TLIVGTMRVNKIGVPLAMFAGAMKMLMPNWIRYPILNIPLLLNLNGLVCGVLAFLFNLQGTP 283

45 Query: 115 ASAGFGISGLIGPINALNLAKGGWCP-VNILLIIIIIFVGAPIVLMNMIFNYLFIKVLKV 171
ASAGFG GL+GPINA L + P V ++ +++ + + +++ + LK+
Sbjct: 284 ASAGFGFIGLVGPINAYRLM--AYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRKLK 339

An alignment of the GAS and GBS proteins is shown below:

50 Identities = 86/101 (85%), Positives = 96/101 (94%)

```
Query: 1 MANVIAKPKIMLPMISSAAILGILGALFNIQGTPASAGFGISGLIGPINALNLAKGGWSV 60
MANVL+KPKIMLPM+ SAA+LG++GA+FNIQGTPASAGFGISGLIGPINALNLAKGGW
```

55 Sbjct: 81 MANVLSKPKIMLPMLCSAAVLGVIGAIFNIQGTPASAGFGISGLIGPINALNLAKGGWCP 140

```
Query: 61 MNMLLIIIIIFVAAPPIIILNFIFNYLFIKVLKIIDPMFYKLDI 101
+N+LIIIIIFV API+LN IFNYLFIKVLK+IDPMFYKLDI
Sbjct: 141 VNILLIIIIIFVGAPIVLMNMIFNYLFIKVLKVIDPMFYKLDI 181
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 393

A DNA sequence (GBSx0426) was identified in *S.agalactiae* <SEQ ID 1277> which encodes the amino acid sequence <SEQ ID 1278>. This protein is predicted to be regulatory protein (pfoS/R). Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
    INTEGRAL Likelihood = -6.58 Transmembrane 148 - 164 ( 145 - 169)
    INTEGRAL Likelihood = -5.26 Transmembrane 33 - 49 ( 25 - 52)
    INTEGRAL Likelihood = -4.73 Transmembrane 70 - 86 ( 62 - 88)
    INTEGRAL Likelihood = -3.45 Transmembrane 124 - 140 ( 122 - 143)
    INTEGRAL Likelihood = -1.33 Transmembrane 96 - 112 ( 96 - 112)

15 ----- Final Results -----
        bacterial membrane --- Certainty=0.3633 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
        bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>
```

20 A related GBS nucleic acid sequence <SEQ ID 9735> which encodes amino acid sequence <SEQ ID 9736> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9697> which encodes amino acid sequence <SEQ ID 9698> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
25 >GP: AAC65034 GB: AE001189 regulatory protein (pfoS/R) [Treponema
      pallidum]
      Identities = 61/158 (38%), Positives = 92/158 (57%)

      Query: 24 KSFIMNVNLNGLALGTIVLIPGAILGELMKALLPMWSGFATLIAATAVATSMMGLVIGIM 83
              + F+M +LNG + G VI L+P AI GEL +AL P+ FA L + +IG +
      Sbjct: 9 RQFMMKILNGSSAGIVIGLVPPAIGELFRALAPLSPLFAALYHVVLPICFSVPALIGTL 68

      Query: 84 VGLNFKNPIQSASLGLAVMFAGGAATFLKGAIMLKGTGDIINMGITAALGVLLIQFLSD 143
              VGL F + + A+L + A G T GA ++ G GD+IN+ + +AL ++L++ L
      Sbjct: 69 VGLQFHCSAPEVATLAFVSIAISGNVTLQNGAWLITGIGDVINVMLISALAIILVRALRG 128

      Query: 144 KTKSFTLIVIPTVTLVVGGVGHVLLPYVKMITTMIGQ 181
              K S T+I +P + ++ GGVG LPVVKMIT +G+
      Sbjct: 129 KLGSLTIIALPVIVAVVAGGVGSFLPYVKMITLFVGR 166
```

40 A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1279> which encodes the amino acid sequence <SEQ ID 1280>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
    INTEGRAL Likelihood = -13.06 Transmembrane 314 - 330 ( 301 - 335)
    INTEGRAL Likelihood = -11.30 Transmembrane 185 - 201 ( 178 - 215)
    INTEGRAL Likelihood = -8.01 Transmembrane 22 - 38 ( 11 - 42)
    INTEGRAL Likelihood = -3.29 Transmembrane 266 - 282 ( 265 - 285)
    INTEGRAL Likelihood = -2.66 Transmembrane 141 - 157 ( 141 - 159)
    INTEGRAL Likelihood = -2.13 Transmembrane 53 - 69 ( 53 - 69)
    INTEGRAL Likelihood = -1.33 Transmembrane 114 - 130 ( 113 - 131)
    INTEGRAL Likelihood = -0.80 Transmembrane 206 - 222 ( 206 - 222)

55 ----- Final Results -----
        bacterial membrane --- Certainty=0.6222 (Affirmative) < succ>
        bacterial outside --- Certainty=0.0000 (Not Clear) < succ>
```

-498-

bacterial cytoplasm --- Certainty=0.0000 (Not Clear) < succ>

The protein has homology with the following sequences in the databases:

5 >GP:AAC65034 GB:AE001189 regulatory protein (pfoS/R) [Treponema
pallidum]
Identities = 137/346 (39%), Positives = 217/346 (62%), Gaps = 14/346 (4%)

 10 Query: 12 FMNKVLAGTAIAIVVALIPNAILATFLKPLLP-NMAAAEFLHIVQVFQFFTPIMAGFLIG 70
 FM K+I G++ IV+ L+P AI + L P + A H+V QF P + G L+G
Sbjct: 11 FMMKILNGSSAGIVIGLVPPIAGELFRALAPLSPLFAALYHVVLPIQFSVPALIGTLVG 70

 15 Query: 71 QQFKFNPMQQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLDINMMITASLAVLA 130
 QF + + + + I SG + G + + GIGD+IN+M+ ++LA++
Sbjct: 71 LQFHCSAPEVATLAFVSVIASG-----NVTLQNGAWLITGIGDVINVMLISALAIIL 122

 20 Query: 131 VKYFGNKFGLSLTIILLPITIGTGVGYIGWKFLPYVSYVTTLIGQGINSFTLQPILMSIL 190
 V+ K GSLTII LP+ + G + G LPYV +T +G+ I +F LQP+LMSIL
Sbjct: 123 VRALRGKLGSLTIIALPVIVAVVAGGVGSFLPYVKMITLFVGRVIATFIALQPLLMSIL 182

 25 Query: 191 IAVAFSLIIIVSPISTVAIGLAIGLNGMAAGAASMGIASTAAVLWATLKVNKSGVPIAIA 250
 +++++FSLII+SP+S+VA+G+A+GL G+A+GAA++G++S A L+ T++VNK GVP+A+
Sbjct: 183 LSMSFSLIIISPVSSAVGIAVGLTGLASGAANIGVSSCAMTLIVGTMRVNKIGVPLAMF 242

 30 Query: 251 LGAMKMMMPNFLKHPIMAIPMVFTAIISSLTVPLFNLVGTTPASSGFGLVGAVGPIAS--L 308
 GAMKM+MPN++++PI+ IP++ + + LFNL GTPAS+GFG +G VGPI + L
Sbjct: 243 AGAMKMLMPNWIRYPIPINIPLLNLNGLVCGVLAWLFBNLQGTPASAGFGFIGLVGPINAYRL 302

 35 Query: 309 AGGSSIL---IIILAWIIVPFAVAFAAHKVSNDILKLYKEDIFVFE 351
 + ++ I+ L + ++ F A+ + D LKLY+ ++F+ E
Sbjct: 303 MAYTPMVRAGILFLVYFVLSFLAAYLIDFILVDRLKLYRRELFIPE 348

An alignment of the GAS and GBS proteins is shown below:

Identities = 65/172 (37%), Positives = 95/172 (54%), Gaps = 9/172 (5%)

35 Query: 19 EKQTTKSFIMNVNLNGLALGTVIVLIPGAILGEIMKALLPMWSGFATLIAATAVATSMMGL 78
 +K+T SF+ VL G A+ V+ LIP AIL +K LLP + A + V +
Sbjct: 5 DKETFSSFMNKVLAGTAIAIVVALIPNAILATFLKPLLPNMAA-AEFLHIVQVFQFFTPI 63

 40 Query: 79 VIGIMVGLNFKNPPIQSASLGLAVMFAGGAATFLK-----GAIMLKGTGDIINMGIT 130
 + G ++G FKFPNP+Q ++G A GA + + G L+G GD+INM IT
Sbjct: 64 MAGFLIGQQFKNPQMQLAVGGAAYIGSGAWAYTEVIQKGVATGTFQLRGIGDLDINMMIT 123

 45 Query: 131 AALGVLLIQFLSDTKSFTLIVIPTVTLLLVGGVGHVLLPYVKMITTMIGQG 182
 A+L VL +++ +K S T+I++P VG +G LPYV +TT+IGQG
Sbjct: 124 ASLAVLAVKYFGNKFGLSLTIILLPITIGTGVGYIGWKFLPYVSYVTTLIGQG 175

A related GBS gene <SEQ ID 8567> and protein <SEQ ID 8568> were also identified. Analysis of this protein sequence reveals the following:

50 Lipop: Possible site: -1 Crend: 10
McG: Discrim Score: -13.49
GvH: Signal Score (-7.5): -5.82
 Possible site: 48
>>> Seems to have no N-terminal signal sequence
55 ALOM program count: 5 value: -6.58 threshold: 0.0
 INTEGRAL Likelihood = -6.58 Transmembrane 148 - 164 (145 - 169)
 INTEGRAL Likelihood = -5.26 Transmembrane 33 - 49 (25 - 52)
 INTEGRAL Likelihood = -4.73 Transmembrane 70 - 86 (62 - 88)
 INTEGRAL Likelihood = -3.45 Transmembrane 124 - 140 (122 - 143)
 INTEGRAL Likelihood = -1.33 Transmembrane 96 - 112 (96 - 112)
60 PERIPHERAL Likelihood = 1.85 51
modified ALOM score: 1.82

 *** Reasoning Step: 3

-500-

Query: 120 EVSLIQDGSRKLQELLAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLA 179
 L++DGS K+QE + G +DIG++ P+ + + T + +V+ SH LA
 Sbjct: 121 TFQLVEDGSIKVQEGVGDSILDIGVVVL PANEDIFHSFTIVKET---LMLVVHFSHLA 176

5 Query: 180 TLPEIELNDLRDYKVASLNEHYMLGEMLPRKCRALGFDPHIVFKHNDWEVLIHSLQDLNA 239
 E +L +L+D E ++L + +C GF PHI+++ + W+ + +
 Sbjct: 177 DEKECQLRELKDEPFIFFREDFVLHNRMTECIKAGFRPHIYYETSQWDFISEMVANLG 236

10 Query: 240 VTIPLSEFESISQVQDLCWVPLKDKNNFYPIGIAYRNDTSFS 281
 + +LP + + +PL D + + I +R D S
 Sbjct: 237 IGLLPERICRGLDPEKVKVPIPLVDPVIPWHLAIIRKDRYLS 278

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1283> which encodes the amino acid sequence <SEQ ID 1284>. Analysis of this protein sequence reveals the following:

15 Possible site: 21
 >>> Seems to have no N-terminal signal sequence

----- Final Results -----

20 bacterial cytoplasm --- Certainty=0.1101(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

An alignment of the GAS and GBS proteins is shown below:

Identities = 125/160 (78%), Positives = 144/160 (89%)

25 Query: 135 LAKGKIDIGLLSFPSTRNDITIEPLQTSTKGYKVSIVMPKSHPLATLPEIELNDLRDYKV 194
 L++GKIDIGLLSF S R DITIE LQTSTKGYKVSIV+ K HFLA P+++L DL+ YK+
 Sbjct: 1 LSQGKIDIGLLSFLSIRKDITIELLQTSTKGYKVSIVLLKQHPLAQHPQLKLKDLKGYKI 60

30 Query: 195 ASLNEHYMLGEMLPRKCRALGFDPHIVFKHNDWEVLIHSLQDLNAVТИЛПСЕФЕСИСОВQ 254
 ASLN+HYMLGEMLPRKCRALGF+P IVFKHNDWEVLIHSL DLN +TILPS+FES++QV
 Sbjct: 61 ASLNДHYMLGEMLPRKCRALGFEPDIVFKHNDWEVLIHSLHDLNLTILPSDFESLNQVD 120

35 Query: 255 DLCWVPLKDKNNFYPIGIAYRNDTSFSPMIEEFLSLLKTN 294
 +L W+PL+DKNNFYPIGIAYR+D SFSP+IEEFLSLLKTN
 Sbjct: 121 NLVWIPLQDKNNFYPIGIAYRDDASFSPVIEEFLSLLKTN 160

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

40 Example 395

A DNA sequence (GBSx0429) was identified in *S.agalactiae* <SEQ ID 1285> which encodes the amino acid sequence <SEQ ID 1286>. Analysis of this protein sequence reveals the following:

Possible site: 47
 >>> Seems to have no N-terminal signal sequence

45 ----- Final Results -----

bacterial cytoplasm --- Certainty=0.1833(Affirmative) < succ>
 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
 bacterial outside --- Certainty=0.0000(Not Clear) < succ>

50 Signal peptide: 1-21

A related GBS nucleic acid sequence <SEQ ID 8569> which encodes amino acid sequence <SEQ ID 8570> was also identified.

55 The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.